

مقدمة قصيرة جدًّا

علم الإحصاء

ديفيد جيه هاند

علم الإحصاء

علم الإحصاء

مقدمة قصيرة جدًّا

تأليف
ديفيد جيه هاند

ترجمة
أحمد شكل

مراجعة
محمد فتحي خضر



علم الإحصاء

ديفيد جيه هاند

Statistics

David J. Hand

الطبعة الأولى ٢٠١٦م
رقم إيداع ١٧٨٠٤ / ٢٠١٥
جميع الحقوق محفوظة للناشر مؤسسة هنداوي للتعليم والثقافة
المشهرة برقم ٨٨٦٢ بتاريخ ٢٦ / ٨ / ٢٠١٢

مؤسسة هنداوي للتعليم والثقافة
إن مؤسسة هنداوي للتعليم والثقافة غير مسئولة عن آراء المؤلف وأفكاره
وإنما يعبِّر الكتاب عن آراء مؤلفه
٥٤ عمارات الفتح، حي السفارات، مدينة نصر ١١٤٧١، القاهرة
جمهورية مصر العربية
تليفون: ٢٠٢ ٢٢٧٠٦٣٥٢ + **فاكس:** ٢٠٢ ٣٥٣٦٥٨٥٣ +
البريد الإلكتروني: hindawi@hindawi.org
الموقع الإلكتروني: http://www.hindawi.org

هاند، ديفيد جيه.
علم الإحصاء: مقدمة قصيرة جدًّا/تأليف ديفيد جيه هاند.
تدمك: ٣ ٣٨٨ ٧٦٨ ٩٧٧ ٩٧٨

١–الإحصاء

أ–العنوان

٣١٠

تصميم الغلاف: محمد الطوبجي.

المحتويات

تمهيد

تمثِّل الأفكار والأساليب الإحصائية أساس كل جوانب الحياة الحديثة تقريبًا. في بعض الأحيان يكون دَوْر الإحصاء واضحًا، ولكنْ في كثير من الأحيان تكون الأفكار والأدوات الإحصائية مختفية في الخلفية. وفي كلتا الحالتين، بسبب الوجود الشامل للأفكار الإحصائية، من الواضح أنه من المُفِيد للغاية أن نمتلك بعض الفهْم لها. والهدف من هذا الكتاب هو تقديم مثل هذا الفهم.

يُعاني الإحصاء من سوء فهْم جوهري مؤسِف يُضلِّل الناس عن طبيعته الأساسية. وهذا الاعتقاد الخاطئ هو أنه يتطلب مهارة حسابية كبيرة مُمِلَّة، وأنه، نتيجة لذلك، مجال جافٌّ ومملٌّ يخلو من الخيال أو الإبداع أو الإثارة. بَيْدَ أن هذه صورة خاطئة تمامًا لمجال علم الإحصاء الحديث؛ إذ إنها مبنية على تصور يرجع تاريخه إلى أكثر من نصف قرن. تحديدًا، تتجاهل هذه الصورة تمامًا حقيقة أن أجهزة الكمبيوتر قد غيرت وجه المجال تمامًا؛ إذ حوَّلَتْه من مجال معتمِد على الحساب إلى نظام قائم على استخدام أدوات برمجية متطورة لسَبْر البيانات بحثًا عن الفهم والتنوير. هذا هو ما يتمحور حوله مجال علم الإحصاء الحديث؛ استخدام الأدوات لمساعدة الإدراك وتوفير كلٍّ من وسائل تسليط الضوء، وسُبُل الفهْم، وأدوات الرصد والتوجيه، ونُظُم المساعدة في عملية صنع القرار. كل هذا — وأكثر — يمثِّل جوانب مجال علم الإحصاء الحديث.

يهدف هذا الكتاب إلى منْح القارئ قدْرًا من الفهْم لمجال علم الإحصاء الحديث. من الواضح أنه في كتاب قصير مثل هذا الكتاب لا أستطيع الخوض في التفاصيل؛ لذا بدلًا من التفصيل، فضَّلتُ إلْقاءَ نظرة عامة على المجال بأسْره، في محاولة للتعبير عن طبيعة الفلسفة والأفكار والأدوات والأساليب الإحصائية. وآمُلُ أنْ يَمنَح الكتابُ القارئَ فهمًا

لكيفية عمل مجال علم الإحصاء الحديث، ومدى أهميته، وأن يُعرِّفه — بالطبع — السبب في أهميته.

يعرض الفصل الأول بعض التعريفات الأساسية، مع توضيحات تهدف للتعريف ببعضٍ من قوة الإحصاء وأهميته وإثارته. ويقدم الفصل الثاني بعضًا من أبسط الأفكار الإحصائية؛ الأفكار التي ربما قابلها القارئ بالفعل، والمعنية بالملخَّصات الأساسية للبيانات. ويحذرنا الفصل الثالث من أن صحة أي استنتاجات نستَقِيها تعتمِد كثيرًا على جودة البيانات الخام، ويوضح أيضًا استراتيجيات لجمع البيانات على نحو أكثر كفاءة. وإذا كانت البيانات إحدى ساقَيِ الإحصاء، فإن ساقها الأخرى هي الاحتمال. ويقدم الفصل الرابع المفاهيم الأساسية للاحتمال. واستنادًا على ساقَيِ البيانات والاحتمالات، يبدأ الإحصاء في الفصل الخامس المشي، مع وصف كيفية استقاء المرء للاستنتاجات والتوصل لاستدلالات من البيانات. ويعرض الفصل السادس لَمْحة خاطفة لبعض الأساليب الإحصائية المهمة، مبيِّنًا كيف أنها تشكِّل جزءًا من شبكة مترابطة من الأفكار والطرق لاستخراج الفهم من البيانات. وأخيرًا، يتناول الفصل السابع بعض الطرق التي أثَّر بها الكمبيوتر على الإحصاء.

أودُّ أنْ أشكر إميلي كينواي، وشيلي شانون، ومارتن كرودر، وقارئًا مجهولًا؛ على التعليق على مسوَّدات هذا الكتاب؛ إذ حسَّنت تعليقاتُهم هذا الكتابَ كثيرًا، وساعدَت على تسوية الغموض في التفسيرات. وبالطبع، أي غموض باقٍ هو خطئي أنا وحدي.

ديفيد جيه هاند

إمبريال كوليدج، لندن

الفصل الأول

علم الإحصاء في كل مكان

ردًّا على أولئك الذين يقولون: «ثمة أكاذيب، وأكاذيب بغيضة، وإحصائيات»، غالبًا ما أقتبس قول فريدريك موستلر: «الكذب بالإحصائيات سهل، ولكنَّ الكذب بدونها أسهل.»

(١) علم الإحصاء الحديث

أريد أن أبدأ بتأكيدٍ ربما يَجِده العديد من القراء مفاجئًا: «علم الإحصاء هو أكثر العلوم إثارة.» وهدفي في هذا الكتاب أنْ أوضح لك أن هذه العبارة صحيحة، وأن أبيِّن لك السبب في صحتها. وآمل أن أبدِّد بعض المفاهيم الخاطئة القديمة حول طبيعة الإحصاء، وإظهار ما يبدو عليه علم الإحصاء الحديث، وكذلك توضيح بعضٍ من قوَّته الهائلة، فضلًا عن انتشاره.

وعلى نحو خاص، أريد في هذا الفصل التمهيدي أن أنقل أمرين؛ أولهما: هو نكهة الثورة التي حدثت في العقود القليلة الماضية؛ فأريد أن أشرح كيف تحول الإحصاء من علم فيكتوري جافٍّ معنيٍّ بالتلاعب اليدوي بأعمدة الأرقام إلى تكنولوجيا حديثة متطورة للغاية تنطوي على استخدام أدوات البرمجيات الأكثر تقدمًا. وأريد توضيح كيف يستخدم إحصائيو اليوم هذه الأدوات لدراسة البيانات بحثًا عن البنيات والأنماط، وكيفية استخدامهم لهذه التكنولوجيا لتقشير طبقات الحيرة والغموض وكشْف الحقائق الموجودة تحتها؛ فعِلم الإحصاء الحديث — على غرار التلسكوبات والمجاهر والأشعة السينية وأجهزة الرادار وأجهزة المسح الطبية — يمكِّننا من رؤية أشياء غير مرئية للعين المجردة؛ فهذا العلم يمكِّننا من الرؤية خلال الضباب والارتباك الموجود في العالَم من حولنا؛ من أجل فهم الواقع الأساسي.

هذا إذَن هو أول شيء أريد أن أوصله خلال هذا الفصل: القوة والإثارة الهائلتان اللتان يضمهما علم الإحصاء الحديث، والمصدر الذي جاء منه، والأشياء التي يَقدِر على فعلها. والشيء الثاني الذي أتمنَّى توصيله هو الوجود الكلي للإحصاء؛ فلا يوجد جانب من جوانب الحياة الحديثة لا يمسه علم الإحصاء. إن الطب الحديث مبنيٌّ على علم الإحصاء؛ فعلى سبيل المثال، وُصفت التجارب العشوائية الخاضعة للضبط بأنها «واحدة من أدوات البحث الأبسط والأقوى والأكثر ثورية.» وفهْم العمليات التي تنتشر الأوبئة من خلالها يمنعها من الفتك بالبشر. تعتمد الحكومة القديرة على التحليل الإحصائي الدقيق للبيانات في وصف الاقتصاد والمجتمع؛ وربما يمثل هذا حجةً للإصرار على أن جميع مَن يكونون في الحكومة ينبغي أن يَدرسوا دورات إلزامية في الإحصاء. والمزارعون وتقنيُّو الغذاء ومراكز التسوق يستخدمون جميعًا الإحصاء على نحو ضمني في تحديدِ ما يزرعونه، وكيفية معالجته، وكيفية تغليفه وتوزيعه. ويحدد الهيدرولوجيون مدى الارتفاع اللازم لبناء حواجز الفيضانات من خلال تحليل إحصائيات الأرصاد الجوية. ويبني المهندسون أنظمة الكمبيوتر باستخدام إحصائيات الموثوقية لضمان عدم تعطلها كثيرًا. وتُبنَى نُظُم مُراقَبَة الحركة الجوية على نماذج إحصائية معقَّدة، بحيث تعمل بشكل لحظي (أي في الزمن الحقيقي). وعلى الرغم من أنك قد لا تدرك ذلك، فإن الأفكار والأدوات الإحصائية كامنةٌ في كل جوانب الحياة الحديثة تقريبًا.

(٢) بعض التعريفات

أحد التعريفات الجيدة لعلم الإحصاء أنه «تكنولوجيا استخراج المعنى من البيانات». ومع ذلك، لا يوجد تعريف مثالي؛ فعلى وجه الخصوص، لا يُشير هذا التعريف إلى المصادفة والاحتمال، اللذين يُعدَّان دعامتين أساسيتين للعديد من تطبيقات الإحصاء؛ ومن ثم ربما يتمثَّل تعريف جيد آخَر في أنه «تكنولوجيا التعامل مع عدم اليقين». ومع هذا، قد تضع تعريفات أخرى، أو تعريفات أكثر دقة، مزيدًا من التركيز على الأدوار التي يلعبها علم الإحصاء. وهكذا يمكننا القول إن علم الإحصاء هو العلم الرئيس «للتنبؤ بالمستقبل» أو «لصنع استنتاجات حول المجهول» أو «لإنتاج ملخصات مناسبة من البيانات». وعند جمع هذه التعريفات معًا فإنها تغطِّي على نحو واسع جوهر هذا المجال، على الرغم من أن التطبيقات المختلفة ستوفِّر تجسيدات مختلفة جدًّا لهذا العلم؛ على سبيل المثال، اتخاذ القرارات والتنبؤ والرصد اللحظي والكشف عن الغش والتعداد السكاني وتحليل

تسلسل الجينات كلها تطبيقات للإحصاء، ومع ذلك ربما تتطلب أساليبَ وأدوات مختلفة للغاية. وثمة شيء تجدر ملاحظته حول هذه التعريفات؛ هو أنني تعمدتُ اختيار كلمة «تكنولوجيا» بدلًا من علم؛ فالتكنولوجيا هي تطبيق للعلم واكتشافاته، وهذا هو ماهية الإحصاء؛ تطبيق فهْمنا لكيفية استخراج المعلومات من البيانات، وفهمنا لعدم اليقين. ومع ذلك، يُشار إلى الإحصاء أحيانًا على أنه علم. في الواقع، إحدى المجلات الإحصائية الأكثر إثارة وتشويقًا تُسمَّى بذلك الاسم فحسب: «العلوم الإحصائية».

وحتى الآن في هذا الكتاب — وعلى وجه الخصوص في الفقرة السابقة — تناولتُ «الإحصاء»، ويوجد شيء آخر سنتناوله في هذا الكتاب هو «الإحصائيات»، والإحصائية هي حقيقة رقمية أو ملخص؛ على سبيل المثال، ملخص للبيانات التي تَصِف بعض السكان؛ ربما حجم السكان أو معدَّل المواليد أو معدَّل الجريمة؛ إذن، يدور هذا الكتاب — من ناحية — حول الحقائق الرقمية الفردية. ولكن بالمعنى الحقيقي للغاية فهو يدور حول ما هو أكثر من ذلك بكثير؛ فهو يدور حول كيفية جمْع ومعالجة وتحليل واستنتاج أشياء من هذه الحقائق الرقمية. وهو يدور حول التكنولوجيا نفسها؛ وهذا يَعنِي أن القارئ الآمِل في أن يَجِدَ جداول أعداد في هذا الكتاب (على سبيل المثال «إحصائيات رياضية») فسوف يُصاب بخيبة أمل. ولكن القارئ الآمِل في التوصل لفهم كيفية اتخاذ الشركات للقرارات، وكيفية اكتشاف علماء الفلك لأنواع جديدة من النجوم، وكيفية تحديد الباحثين في مجال الطب للجينات المرتبطة بمرض معين، وكيفية اتخاذ البنوك قرارًا بمنح أو عدم منح شخصٍ ما بطاقةَ ائتمان، وكيفية تحديد شركات التأمين تكلفة القسط، وكيفية بناء مرشِّحات البريد المزعِج التي تمنع الإعلانات المزعِجة من الوصول إلى صندوق بريدك الإلكتروني، وما إلى ذلك؛ فإنه سوف يجد مأربه.

كل ما سبق يبيِّن الفارق بين المُسمَّيَيْن «الإحصاء» و«الإحصائيات»؛ فالإحصاء هو العلم الأساسي الشامل، أما الإحصائيات فيُقصَد بها الحقائق الرقمية أو الملخصات المندرجة تحت المظلَّة الكبرى لعلم الإحصاء.

استخدمتُ في تعريفي الأول كلمة «البيانات». وكلمة «بيانات» في الإنجليزية Data مشتقة من الكلمة اللاتينية datum بمعنى «شيء مُعطًى» المشتقة من dare بمعنى «يعطي». عادة ما تكون البيانات أرقامًا؛ نتائجَ قياساتٍ أو حساباتٍ أو غيرها من العمليات. ويمكن النظر لمثل هذه البيانات على أنها تقدِّم تمثيلًا مبسطًا لما ندرسه. فإذا كنَّا مهتمِّين بأطفال المدارس، وبخاصة قدرتهم الأكاديمية ومدى ملاءمتهم لأنواع

المِهَن المختلفة، ربما نختار دراسة الأرقام التي تصف نتائجهم في مختلف الاختبارات والامتحانات.

وربما تمنحنا هذه الأرقام إشارة حيال قدراتهم وميولهم. باعتراف الجميع، لن يكون هذا التمثيل مثاليًّا؛ فربما تُشِير الدرجة المنخفضة ببساطةٍ إلى أن شخصًا ما كان يشعر بالمرض أثناء الامتحان. وعبارة «لم يحضر» لا تُخبِرنا بالكثير عن قدرة الطفل، ولكن تخبرنا فحسب أنه لم يَخُضْ الامتحان. سأتحدث بشكل أكثر استفاضة عن «جودة البيانات» في وقت لاحق، وهي مهمة بسبب المبدأ العام (الذي ينطبق على جميع جوانب الحياة، وليس فقط في الإحصائيات) القاضي بأنه إذا كانت المادة الخام التي تعمل عليها رديئة، فإن النتائج ستكون رديئة. يستطيع الإحصائيون فهْم أشياء كثيرة مذهلة من الأرقام، لكنهم لا يمكن أن يصنعوا المعجزات.

بطبيعة الحال، يبدو أن حالات كثيرة لا تُنتِج بيانات رقمية مباشرة؛ فالكثير من البيانات الخام قد تكون في شكل صور أو كلمات أو حتى أشياء مثل إشارات إلكترونية أو صوتية؛ ومن ثم فإن صور الأقمار الصناعية للمحاصيل أو تغطية الغابات المَطِيرة، والأوصاف اللفظية للآثار الجانبية التي تَحدث عند تناول الدواء، والأصوات الملفوظة عند التحدث؛ لا تأخذ مَظهَر الأرقام. ومع ذلك، يُظهِر الفحص الدقيق أنه عندما تُقاس هذه الأشياء وتُسجَّل، فإنها تُترجَم إلى تمثيلات رقمية أو إلى تمثيلات يمكن أن تُترجَم بعد ذلك إلى أرقام؛ على سبيل المثال، صور الأقمار الصناعية والصور الأخرى تُمثَّل بملايين العناصر الصغيرة التي تُسمَّى وحدات البكسل، وكلٌّ منها يوصف من حيث الشدة (الرقمية) للألوان المختلفة التي تشكِّلها. ويمكن معالَجة النص في صورة تعداد للكلمات أو مقاييس للتشابه بين الكلمات والعبارات؛ وهذا هو نوع التمثيل المستخدم من قِبَل محرِّكات البحث على شبكة الإنترنت مثل جوجل. وتُمثَّل الكلمات المنطوقة من خلال الكثافات الرقمية للأشكال الموجية التي تشكِّل الأجزاء المفردة من الكلام. وعلى نحو عام، رغم أنه ليست جميع البيانات أرقامًا، فإن معظم البيانات تُترجَم إلى شكل رقمي في مرحلةٍ ما. ومعظم الإحصائيات تتعامل مع البيانات الرقمية.

(٣) أكاذيب، أكاذيب بغيضة، ووضع الأمور في نصابها

نُسبت عبارة «ثمة أكاذيب، وأكاذيب بغيضة، وإحصائيات» — المذكورة في بداية هذا الفصل — على وجوه مختلفة إلى مارك توين وبنيامين دزرائيلي، وغيرهما. كما وَرَدَ على

لسان العديد من الأشخاص تصريحات مماثلة؛ منها: «على غرار الأحلام، الإحصائيات هي شكل من أشكال تحقيق الرغبات» (جون بودريار، في كتاب «ذكريات جميلة»، الفصل الرابع)، و«... عبادة الإحصائيات أدَّتْ على نحو خاص إلى نتيجة مؤسِفة تمثَّلتْ في جعل مهمة الكاذب الصرف أسهل بكثير» (توم بورنام، في كتاب «قاموس التضليل»)، و«الإحصائيات هي «خُزَعْبِلات» مدعومة بالأرقام» (أودري هابيرا وريتشارد رونيون، في كتاب «الإحصائيات العامة»)، و«الإجراءات القانونية مثل الإحصائيات؛ إذا تلاعبتَ بها، يمكنك أن تثبت أيَّ شيء» (آرثر هيلي، في رواية «المطار»)، وما إلى ذلك.

من الواضح أنه يوجد كثير من الشك حيال الإحصائيات، وربما نتساءل أيضًا ما إذا كان هناك عنصر خوف من هذا المجال. من المؤكد أن الإحصائي غالبًا ما يلعب دور شخص يجب عليه توخِّي الحَذَر، وربما حتى يكون حاملَ الأخبار السيئة. والإحصائيون العاملون في البيئات البحثية — على سبيل المثال في كليات الطب أو السياقات الاجتماعية — ربما يكون عليهم شرح أن البيانات غير كافية للإجابة عن سؤال معين، أو أن الجواب ببساطة ليس ما أراد الباحث سَمَاعَه، وربما يكون هذا أمرًا مؤسفًا من وجهة نظر الباحث، ولكن ليس من الإنصاف إلقاء اللوم على حامل الرسالة الإحصائية.

في كثير من الحالات، تتولَّد الشكوك بسبب أولئك الذين يختارون الإحصائيات انتقائيًّا. فإذا كان هناك أكثر من طريقة لتلخيص مجموعة من البيانات، وتتبع كلٌّ منها بالنظر في جوانب مختلفة قليلًا، فإن الأشخاص المختلفين حينَها يمكن أن يختاروا التركيز على ملخصات مختلفة. وثمة مثال محدد في إحصائيات الجريمة؛ ففي بريطانيا، ربما يُعَدُّ أهم مصدر لإحصائيات الجريمة هو «استقصاء الجريمة البريطانية»، وهذا الاستقصاء يُقدِّر مستوى الجريمة عن طريق سؤال عيِّنة من الناس مباشرة عن الجرائم التي وقعوا ضحايا لها خلال العام الماضي. في المقابل، فإن سلسلة «إحصائيات الجرائم المسجَّلة» تشمل جميع الجرائم المُبلَّغ عنها إلى وزارة الداخلية والتي سجَّلتْها الشرطة. وبطبيعتها، لا تشمل هذه الإحصائيات بعض الجرائم البسيطة، وأهم من ذلك بطبيعة الحال أنها تستثني الجرائم التي لم تُبلَّغ عنها الشرطة في المقام الأول. وبوجود مثل هذه الاختلافات، ليس من المستغرَب أن الأرقام يمكن أن تختلف بين مجموعتَي الإحصائيات، لدرجة أن فئات معينة من الجرائم ربما تبدو آخِذة في التناقص على مر الزمن وفقًا لإحدى مجموعتَي الأرقام فيما تكون آخذة في التزايد وفقًا للمجموعة الأخرى.

أرقام إحصائيات الجريمة توضح أيضًا سببًا محتملًا آخر للتشكك في الإحصائيات؛ فعند استخدام مقياس معين كمؤشر لأداء نظامٍ ما، ربما يختار الأشخاص استهداف هذا

المقياس، فيُحسنون قيمته ولكنْ على حساب جوانب أخرى من النظام؛ ومن ثم يتحسن المقياس المختار على نحو غير متكافئ، ويصبح عديم الفائدة كمقياس لأداء النظام؛ على سبيل المثال، يمكن للشرطة أن تقلل من معدل سرقة المتاجر من خلال تركيز كل مواردها على تلك الجريمة، على حساب السماح بزيادة أنواع أخرى من الجريمة؛ ونتيجة لذلك، فإن معدل سرقة المتاجر يصبح عديم الفائدة كمؤشر على معدل الجريمة. وقد سُمِّيت هذه الظاهرة باسم «قانون جودهارت»، تيمُّنًا بتشارلز جودهارت، وهو كبير مستشارين سابقًا في «مصرف إنجلترا».

الهدف من كل ذلك هو أن المشكلة لا تكمن في الإحصائيات في حد ذاتها، ولكن في استخدام تلك الإحصائيات، وسوء فهم كيفية إنتاج الإحصائيات، وما تعنيه الإحصائيات حقًّا. لعل من الطبيعي تمامًا أن نكون متشكِّكين حيال الأشياء التي لا نفهمها، والحل هو إزالة سوء الفهم.

مع ذلك، ثمة سبب آخر للتشكك ينشأ أساسًا نتيجة لطبيعة التقدم العلمي؛ ومن ثَمَّ، ربما نقرأ في يوم من الأيام في صحيفةٍ ما عن دراسة علمية تبيِّن أن نوعًا معينًا من الطعام ضارٌّ لنا، وفي اليوم التالي تُشِير إلى أنه مُفِيد. بطبيعة الحال يولِّد ذلك الْتِباسًا؛ أيْ شعورًا بأن العلماء لا يعرفون الجواب، وربما أنه لا يمكن الوثوق بهم. وحتمًا مثل هذه التحقيقات العلمية تَستخدم التحليلات الإحصائية على نحو مكثَّف؛ ومن ثم فإن بعضًا من هذه الشكوك ينتقل إلى الإحصائيات. ولكن جوهر التقدم العلمي هو تحقيق اكتشافات جديدة تغيِّر فهمنا؛ فرغم أننا كنَّا نظن في الماضي أن الدهون الغذائية ضارة لنا، فقد دفعَنا مزيد من الدراسات إلى إدراك أنه يوجد أنواع مختلفة من الدهون؛ بعضها مفيد وبعضها ضار. إن الصورة أكثر تعقيدًا مما كنَّا نعتقد في البداية؛ لذلك ليس من المستغرَب أن تؤدِّي الدراسات الأولية إلى استنتاجات تبدو متضاربة ومتناقضة.

والسبب الرابع للتشكك ينشأ من سوء فهم أوليٍّ لمبادئ الإحصاء. وكتمرين، ربما يحاول القارئ أن يحدد ما هو مثير للشكوك في كلٍّ من العبارات التالية (الأجوبة موجودة في التعليقات الختامية في آخِر الكتاب):

(١) نقرأ في تقريرٍ ما أن التشخيص المبكِّر للمرض يؤدي إلى التمتع بمعدلات عمرية أطول؛ لذلك فإن برامج الفحص مفيدة.

(٢) قيل لنا إن السعر المُعلَن خُفِّض بالفعل بنسبة خصم ٢٥٪ للعملاء المؤهَّلين، ولكننا لسنا مؤهَّلين؛ لذلك علينا دفع ٢٥٪ أكثر من السعر المُعلَن.

(٣) نسمع تنبؤًا بأن متوسط العمر المتوقَّع سوف يصل إلى ١٥٠ عامًا في القرن المُقبِل، استنادًا إلى استقراء بسيط من الزيادات على مدى السنوات المائة الماضية.

(٤) قيل لنا: «منذ عام ١٩٥٠، تَضاعَف كلَّ عام عددُ الأطفال الأمريكيين الذين تعرضوا لحادث إطلاق نار.»

أحيانًا لا يكون سوء الفهم أوليًّا للغاية، أو على الأقل، ينشأ عن مفاهيم إحصائية عميقة نسبيًّا. سيكون مستغربًا ألَّا يوجد بعض الأفكار العميقة المناقضة للبديهة في الإحصاء بعد أكثر من قرن من التطور. وتتمثل إحدى هذه الأفكار فيما يُعرَف باسم «مُغالَطة المدعي»، وتصف الخلط بين احتمال أن شيئًا ما سوف يكون صحيحًا (على سبيل المثال، المتهم مذنب) إذا كان لديك بعض الأدلة (على سبيل المثال، قفازات المدعَى عليه في مسرح الجريمة)، مع احتمال العثور على هذا الدليل إذا كنتَ تفترض أن المتهم مذنب. وهذا خلط شائع — ليس في المحاكم فحسب — وسوف نتناوله على نحو أوثق في وقت لاحق.

إذا كان هناك شك وعدم ثقة في الإحصائيات، فمِن الواضح أن اللوم لا يَقَع على الإحصائيات أو كيفية حسابها، وإنما يقع على طريقة استخدام تلك الإحصائيات. وليس من العدل إلْقاء اللوم على العلم، أو الإحصائي الذي يَستخرج المعنى من البيانات؛ بل إن اللوم يقع على أولئك الذين لا يفهمون ما تقوله الأرقام، أو الذين يتعمَّدون إساءة استخدام النتائج؛ فنحن لا نلوم البندقية على قتْل أحدهم، بل الشخص الذي أطلق الرصاص من البندقية هو المَلُوم.

(٤) البيانات

رأينا أن البيانات هي المادة الخام التي بُنِي عليها الإحصاء، وكذلك هي المادة الخام التي تُحسب منها الإحصائيات الفردية نفسُها، وأن هذه البيانات عادةً ما تكون أرقامًا. ومع ذلك، فإن البيانات في الواقع أكثر من مجرد أرقام. ولكي تكون مفيدة — أيْ تمكِّننا من القيام ببعض التحليلات الإحصائية ذات المغزى — يجب أن ترتبط هذه الأرقام بمعنًى؛ فعلى سبيل المثال، نحن بحاجة إلى معرفةِ ما «تَقِيسه» القياسات، وما تم عَدُّه عندما يُعرض علينا تعداد. ولتحقيق نتائج صحيحة ودقيقة عندما نقوم بتنفيذ تحليل إحصائي، نحتاج أيضًا أن نعرف شيئًا عن كيفية الحصول على هذه القِيَم. هل أجاب جميع مَن سألناهم

على الاستبيان، أم أجاب بعض الأشخاص فحسب؟ وإذا أجاب بعض الأشخاص فحسب، فهل هم يمثِّلون المجموعة التي نودُّ أن نُدلِيَ ببيان حولَها على نحو ملائم أم إن العينة مشوَّهة بطريقةٍ ما؟ هل، على سبيل المثال، تَستبعِد عيِّنتنا الشبابَ على نحو غير متكافئ؟ وبالمثل، فإننا بحاجة إلى معرفةِ ما إذا انسحب مرضى من التجارب السريرية، وما إذا كانت البيانات مُحدَّثة أم لا. ونحتاج إلى معرفة ما إذا كانت أداة القياس موثوقًا بها أم لا، أو هل كانت لديها قيمة قصوى تُسجَّل عندما تكون القيمة الحقيقية مرتفِعة على نحو مفرط. هل لنا أن نفترض أن معدل النبض الذي سجَّلتْه الممرضة دقيق أم إنه قِيمة تقريبية فحسب؟ ثمة عدد لا حصر له من مثل هذه الأسئلة التي يمكن طرحه، ونحتاج إلى أن نكون متنبِّهين لتلك الأسئلة التي يمكن أن تؤثِّر على النتائج التي نستخلصها. وإذا لم نفعل ذلك، فستصبح الشكوك من النوع المذكور آنفًا مشروعة تمامًا.

تتمثَّل إحدى طرق النظر إلى البيانات في اعتبارها «أدلة»؛ فبدون بيانات، تصبح أفكارنا ونظرياتنا حيال العالم محض تكهنات. وتوفِّر البيانات معرفةً أساسيةً تربط أفكارنا ونظرياتنا بالواقع، وتسمح لنا بالتحقق من صحة فهمنا واختباره. بعد ذلك تُستخدَم الأساليب الإحصائية لمقارنة البيانات مع أفكارنا ونظرياتنا، لنرى مدى توافق بعضها مع بعض. وسوء التوافق يدفعنا إلى التفكير مرة أخرى وإعادة تقييم أفكارنا وإعادة صياغتها لكي تتطابق على نحو أفضل مع الواقع المرصود. ولكن ربما يجدر وضع ملاحظة تحذيرية هنا؛ وهي أن سوء التوافق يمكن أيضًا أن يكون ناتجًا عن سوء جودة البيانات. يجب أن نكون منتبهين لهذا الاحتمال؛ فربما تكون نظرياتنا سليمة ولكن قد تكون أدوات القياس مَعيبةً بطريقةٍ ما. ومع ذلك، فالتطابق الجيد بين البيانات المرصودة وما تقوله نظرياتنا عمَّا ينبغي أن تكون عليه البيانات يؤكد عمومًا على أننا على الطريق الصحيح. وذلك يؤكد على أن أفكارنا تعكس حقًّا حقيقةَ ما يجري.

يستتبع ذلك ضمنًا أنه لكي تكون أفكارنا ونظرياتنا ذات مغزًى، يجب أن تُسفِر عن توقعات يمكن مقارنتها مع البيانات الموجودة لدينا. فإذا لم تُخبِرنا النظريات بما ينبغي أن نتوقع ملاحظته، أو إذا كانت التوقعات عامة للغاية بحيث إن أي بيانات سوف تتوافق مع نظرياتنا، فإنها لن تكون ذات فائدة كبيرة؛ فأي بيانات ستتوافق معها. وقد انتُقد التحليل النفسي والتنجيم على هذه الأسس.

كما تسمح البيانات لنا بتحسس طريقنا عبر العالم المعقَّد؛ باتخاذ قرارات حول أفضل الإجراءات التي يجب القيام بها؛ فنحن نأخذ قياساتنا، ونحسب المجاميع الكلية،

ونستخدم الأساليب الإحصائية لاستخراج المعلومات من هذه البيانات لوصف الكيفية التي يسير بها العالم وما علينا أن نفعل لجعله يسير على النحو الذي نريد. وهذه المبادئ توضحها أشياء مثل الطيار الآلي في الطائرة، وأنظمة المِلاحة بالأقمار الصناعية في السيارات، والمؤشرات الاقتصادية مثل معدل التضخم والناتج المحلي الإجمالي، ومراقبة المرضى في وحدات العناية المركزة، وتقييم السياسات الاجتماعية المعقدة.

ونظرًا للدور الأساسي الذي تلعبه البيانات بوصفها الرابط بين ملاحظاتنا للعالم من حولنا وبين أفكارنا وفهمنا لهذا العالم، فإنه ليس من قَبِيل المبالغة أن نَصِفَ البيانات — وتكنولوجيا استخراج المعنى منها — باعتبارها حجر الأساس للحضارة الحديثة. وهذا هو السبب في أنني استخدمتُ العنوان الفرعي «كيف تتحكم البيانات في عالمنا؟» لكتابي «توليد المعلومات» (انظر قسم القراءات الإضافية).

(٥) علم الإحصاء الأعظم

على الرغم من أن جذور مجال الإحصاء يمكن تتبعها لزمن بعيد للغاية، فإن مجال الإحصاء نفسه في الحقيقة يبلغ من العمر بضعة قرون فحسب. تأسست الجمعية الإحصائية الملكية في عام ١٨٣٤، والجمعية الإحصائية الأمريكية في عام ١٨٣٩، في حين أنه لم يُنشأ قسم للإحصاء في أي جامعة في العالم حتى عام ١٩١١، حين حدث ذلك في يونيفرسيتي كوليدج في لندن. تضمن مجال الإحصاء المبكر عدة فروع، تجمَّعَت في نهاية المطاف لتصبح علم الإحصاء الحديث. تَمثَّل أحد هذه الفروع في فهم الاحتمالات، وهو أمر يعود تاريخه إلى منتصف القرن السابع عشر، ونبع جزئيًّا من الأسئلة المتعلِّقة بالمقامرة. وتَمثَّل آخَر في إدراك أن القياسات نادرًا ما تكون خالية من الأخطاء، ولذلك وُجدت حاجة إلى بعض التحليل لاستخراج معنًى معقول منها. وفي السنوات الأولى، كان هذا مهمًّا، خصوصًا في علم الفلك. ولكن كان يوجد فرع آخر وهو الاستخدام التدريجي للبيانات الإحصائية لتمكين الحكومات من إدارة بلدانها. وفي الواقع، هذا الاستخدام هو الذي أدَّى إلى ظهور كلمة Statistics بمعنى «إحصائيات»؛ فهي بيانات عن الدولة State. وتمتلك كل الدول المتقدمة الآن مكاتب إحصاء وطنية خاصة بها.

مرَّ علم الإحصاء، خلال تطوره، بعدة مراحل. تميَّزت المرحلة الأولى — التي امتدَّت حتى نهاية القرن التاسع عشر تقريبًا — بالاستكشافات العشوائية للبيانات. ثم شهد النصف الأول من القرن العشرين اكتساب الإحصاء للصبغة الرياضية، لدرجة أن الكثيرين

رأَوْها فرعًا من الرياضيات (إنها تتعامل مع الأرقام، أليس كذلك؟) وبالفعل، لا يزال الإحصائيون في الجامعة غالبًا ما يدرسون الإحصاء داخل أقسام الرياضيات. شَهِد النصف الثاني من القرن العشرين ظهور الكمبيوتر، وكان هذا التغيير هو الذي ارتَقَى بالإحصاء من كونها عملًا صعبًا إلى عمل مُمْتِع؛ فقد أزال الكمبيوتر الحاجة لامتلاك ممارسي الإحصاء لمهارات حسابية خاصة، فلم يعودوا بحاجة لقضاء ساعات طويلة في معالجة الأرقام. وهذا مماثل للتغيير من الحاجة إلى المشي إلى كل مكان للقدرة على قيادة السيارة؛ فالرحلات التي كانت تستغرق في السابق أيامًا أصبحتِ الآن تستغرق دقائق، والرحلات التي كانت طويلة للغاية لدرجة تمنَع التفكير فيها أصبحت الآن ممكنة.

شهد النصف الثاني من القرن العشرين أيضًا ظهور مدارس أخرى لتحليل البيانات، لا تعود أصولها لعلم الإحصاء الكلاسيكي ولكنْ لمجالات أخرى، خاصة علوم الكمبيوتر. وتشمل هذه المدارس التعلم الآلي والتعرف على الأنماط والتنقيب عن البيانات. وبينما تطورت هذه التخصصات الأخرى، كانت تحدث في بعض الأحيان توترات بين هذه المدارس المختلفة والإحصاء. ومع ذلك، فالحقيقة هي أن وجهات النظر المتفاوتة التي تقدِّمها هذه المدارس المختلفة ساهمت جميعها بشيء في تحليل البيانات، إلى حدِّ أن الإحصائيين الجدد في الوقت الحالي يختارون بحُرية من الأدوات التي توفِّرها جميع هذه المجالات. وسأذكر بعض هذه الأدوات في وقت لاحق. بوضع هذا في الاعتبار، سوف أتبنَّى في هذا الكتاب تعريفًا واسعًا للإحصاء، مهتديًا بتعريف «علم الإحصاء الأعظم» الذي قدَّمه الإحصائي البارز جون تشامبرز، الذي قال: «يمكن تعريف علم الإحصاء الأعظم ببساطة — وإنْ كان على نحو غير مُحكَم — بأنه كل ما يتعلق «بالتعلم من البيانات»، من التخطيط أو الجمع الأول حتى العرض أو التقرير الأخير.» أما محاولة وضْع حدود بين تخصصات تحليل البيانات المختلفة، فهي عملية غير مُجدِية ولا طائل من ورائها.

إذن، علم الإحصاء الحديث لا يَدُور حول الحساب، وإنما يدور حول «الاستقصاء»، بل إن البعض وَصَفَ علم الإحصاء بأنه «تطبيق الأسلوب العلمي». ومع أننا ما زلنا نجد في كثير من الأحيان أن العديد من الإحصائيين يعملون انطلاقًا من أقسام الرياضيات في الجامعات كما أشرتُ آنفًا، فإننا نَجِدهم أيضًا في كليات الطب وأقسام العلوم الاجتماعية، بما في ذلك الاقتصاد والعديد من الأقسام الأخرى التي تتراوح بين الهندسة إلى علم النفس. وفي خارج الجامعات، تعمل أعداد كبيرة في الحكومة والصناعة، وفي القطاع الدوائي، والتسويق، والاتصالات، والخدمات المصرفية، ومجموعة كبيرة من المجالات الأخرى، فجميع

المُديرين يعتمدون على المهارات الإحصائية لمساعدتهم في تفسير البيانات التي تَصِف أقسامهم وشركاتهم وإنتاجهم والموظفين وما إلى ذلك. لا يستخدم هؤلاء الأشخاص الرموز والصيغ الرياضية، ولكن يستخدمون الأدوات والأساليب الإحصائية لاكتساب المعرفة والفهم من الأدلة؛ أي البيانات. وللقيام بذلك، فإنهم يحتاجون إلى دراسة مجموعة واسعة من الأمور غير الرياضية في جوهرها؛ مثل جودة البيانات، وشكلها وكيفية جمعها، وتحديد المشكلة، وتحديد الهدف الأكبر للتحليل (الفهم والتنبؤ والقرار، وما إلى ذلك)، مع تحديد مقدار عدم اليقين المرتبط بالنتائج، ومجموعة من الأمور الأخرى.

كما آمل أن يكون قد اتضح مما سبق، فإن علم الإحصاء كلِّيُّ الوجود؛ إذ يتخلل جميع مناحي الحياة. وقد كان لذلك تأثير متبادَل على تطور علم الإحصاء نفسه؛ فبينما طُبقت الأساليب الإحصائية في مجالات جديدة، أدت المشاكل والمتطلبات والخصائص المعينة لتلك المجالات إلى تطوير أساليب وأدوات إحصائية جديدة. وبعد ذلك، بمجرد أن طُورت هذه الأساليب والأدوات الجديدة، انتشرت ووجدت تطبيقات لها في مجالات أخرى.

(٦) بعض الأمثلة

مثال ١: فَلْترة البريد المزعج

«البريد المزعج» هو مصطلح يُستخدم لوصف رسائل البريد الإلكتروني غير المرغوب فيها المُرسَلة تلقائيًّا إلى العديد من المتسلِّمين؛ عادةً ما يَصِل عددهم إلى ملايين المتسلِّمين. هذه الرسائل رسائل دعائية، وغالبًا ما تكون مُزعِجة، وربما تكون واجهات لمُحتالِين. وهي تشمل أشياء مثل عروض دمج الديون، وخطط الثراء السريع، والأدوية التي لا تُصرَف إلا بوصفة طبية، ونصائح حول سوق الأسهم، وأدوات جنسية غريبة. والمبدأ الأساسي في هذه الرسائل هو أنه إذا راسلتَ عددًا كافيًا من الناس، من المحتمل أن يُصبِح بعضهم مهتمًّا — أو ينخدع — بعرضك. وما لم تكن الرسائل آتية من منظمات طُلب منها على وجه التحديد معلومات، فإن معظمها لن يكون مثيرًا للاهتمام، ولن يرغب أحد في تضييع وقته في قراءتها وحذفها. وهو ما يقودنا إلى مرشحات البريد المزعج؛ وهي برامج حاسوبية تفحص تلقائيًّا رسائل البريد الإلكتروني الواردة وتحدد الرسائل التي من المحتمل أن تكون غير مرغوب فيها. ويمكن برمجة المرشحات بحيث يَحذف البرنامج الرسائل غير المرغوب فيها تلقائيًّا، أو يرسلها إلى مجلد تخزين للفحص لاحقًا، أو يتخذ بعض الإجراءات

الأخرى المناسبة. توجد تقديرات مختلفة لكمية البريد المزعج التي تُرسل، ولكن في وقت كتابة هذا الكتاب، يُشير أحد التقديرات إلى أنه ترسل أكثر من ٩٠ مليار رسالة من البريد غير المرغوب فيه كل يوم؛ وبما أن هذا العدد يرتفع ارتفاعًا كبيرًا كل شهر، فمن المرجح أن يكون أكبر بكثير في وقت قراءتك لهذا الكتاب.

ثمة تقنيات عديدة لمنع البريد غير المرغوب فيه. تتحقق بعض الطرق البسيطة للغاية فحسب من وجود كلمات أساسية في الرسالة؛ على سبيل المثال، إذا كانت رسالة تتضمن كلمة viagra «فياجرا»، ربما تُحظر. ومع ذلك، فإن إحدى خصائص رصد البريد المزعج هي أنها تشبه سباق التسلح؛ فبمجرد أن يدرك المسئولون عن الرسائل أن رسائلهم حُظرت بطريقة معينة، يسعَوْن إلى أساليب للالتفاف حول هذه الطريقة؛ على سبيل المثال، ربما يتعمدون كتابة viagra على نحو خاطئ في صورة v1agra أوv-iagra؛ بحيث يمكنك التعرف عليها ولكن دون أن يتمكن البرنامج التلقائي من التعرف عليها.

تستند أدوات رصد البريد غير المرغوب فيه الأكثر تطورًا على نماذج إحصائية للمحتوى الكلامي لرسائل البريد غير المرغوب فيه؛ فعلى سبيل المثال، ربما تَستخدم تقديرات لاحتمالات وجود كلمات معينة أو مجموعات من الكلمات التي تَظهَر في رسائل البريد غير المرغوب فيه. وبعد ذلك، تُصبِح الرسالة التي تحتوي على الكثير من الكلمات العالية الاحتمال موضعَ شَكٍّ. وتبني الأدوات الأكثر تطورًا نماذج لاحتمالية أن كلمة واحدة ستتبع كلمة أخرى في تسلسل؛ ومن ثَمَّ تتمكن من رصد العبارات ومجموعات الكلمات المشبوهة. علاوة على ذلك، تستخدم أساليب أخرى نماذج إحصائية للصور لرصد أشياء مثل لون البشرة في الصورة المرسلة عبر البريد الإلكتروني.

مثال ٢: قضية سالي كلارك

في عام ١٩٩٩، خضعت سالي كلارك — وهي محامية بريطانية شابة — للمحاكمة وأُدينت وحُكم عليها بالسجن مدى الحياة لقتْلها طفلَيْها. توفي طفلها الأول في عام ١٩٩٦، عن عمر يبلغ ١١ أسبوعًا، ومات طفلها الثاني في عام ١٩٩٨، عن عمر يبلغ ٨ أسابيع. واعتمد الحكم على ما أصبح نموذجًا لسوء فهم واستخدام الإحصائيات، عندما ادَّعى طبيب الأطفال السير روي مِدو، في دوره كشاهد خبير لصالح الادِّعاء، أن احتمالية الموت المفاجئ لطفلين كانت ١ من بين ٧٣ مليون حالة. وقد حصل على هذا الرقم ببساطة عن طريق ضرب احتمالية حالتَي الوفاة معًا على نحو منفصل. وبقيامه بذلك، ولجهله

بأساسيات الإحصاء، تجاهَل تمامًا حقيقة أن حدوث واحدة من حالات الوفاة تلك في أي أسرة من المرجح أن يعني ارتفاع احتمالية حدوث وفاة أخرى.

تُبيِّن دراسة البيانات السابقة أن احتمال تعرُّض أي طفل مختار عشوائيًّا للموت المفاجئ في أسرة مثل أسرة كلارك يبلغ حوالي ١ / ٨٥٠٠. وإذا افترضنا بالتبعية أن وقوع حالة وفاة مثل هذه لا يُغير احتمال وقوع حالة أخرى، فإن فرصة وقوع حالتين من هذه الوفيات في الأسرة نفسها ستكون ١ / ٨٥٠٠ مضروبًا في ١ / ٨٥٠٠؛ أي واحدًا من ٧٣ مليونًا. بَيْدَ أن هذا الافتراض جريء، ويُشير التحليل الإحصائي الدقيق للبيانات السابقة إلى أنه في الواقع تزداد فرصة حدوث موت مفاجئ ثانٍ كثيرًا عند وقوع حالة مماثلة قبل ذلك بالفعل. وفي الواقع، تشير الحسابات إلى أن العديد من حالات الوفاة المتعدِّدة تلك ينبغي أن يُتوقَّع حدوثها كل عام في دولة بحجم المملكة المتحدة. ويقول الموقع الإلكتروني لمؤسسة دراسة أسباب موت الأطفال: «من النادر جدًّا حدوث الموت المفاجئ مرتين في الأسرة نفسها، على الرغم من أن اضطرابًا وراثيًّا في بعض الأحيان — مثل وجود خلل أيضي — قد يسبب موتَ أكثرَ من رضيع على نحو غير متوقع.»

في قضية سالي كلارك، كان يوجد مزيد من الأدِلَّة التي تُشِير إلى براءتها، وفي النهاية أصبح من الواضح أن ابنَها الثاني كان يُعانِي عَدْوَى بكتيرية معروفًا أنها تسبِّب موت الرضيع المفاجئ. وأطلق سراح السيدة كلارك بعد ذلك في الاستئناف في عام ٢٠٠٣. ومن المأساوي أنها توفيت في مارس من عام ٢٠٠٧ عن عمر يبلغ ٤٢ عامًا فحسب. ويوجد مزيد من التفاصيل عن سوء الفهم الرهيب وسوء استخدام الإحصائيات في مقال ممتاز كتبتْه هيلين جويس على الموقع المذكور في قسم القراءات الإضافية في نهاية هذا الكتاب.

مثال ٣: عناقيد النجوم

مع ازدياد قدرتنا على سَبْر المزيد والمزيد من أغوار الكون، أصبح من الواضح أن الأجرام السماوية تميل إلى التجمُّع معًا، وتفعل ذلك بطريقة هرمية؛ حيث تشكل النجوم عناقيد، وعناقيد النجوم نفسها تشكل عناقيد على مستوًى أعلى، وهذه العناقيد الأعلى تتجمع بدَوْرها في عناقيد أكبر. وعلى وجه التحديد، مجرَّتنا — والتي هي عنقود من النجوم — جزء من «المجموعة المحلية» المكونة من حوالي ثلاثين مجرَّة، وهذه المجموعة بدورها جزء من «العنقود المجرِّي المحلي الهائل». على النطاق الأوسع، يبدو الكون بالأحرى مثل الرغوة، مع وجود خيوط تتكون من عناقيد مجرية فائقة واقعة على حواف مساحات

فارغة شاسعة. ولكن كيف اكتُشف كل هذا؟ فحتى لو استخدمنا تلسكوبات قوية للنظر خارج الأرض، فإننا نرى ببساطة سماءً مليئة بالنجوم. والجواب هو أن استنتاج وجود هذا الهيكل العنقودي — بل واكتشافه في المقام الأول — تَطلَّب تقنيات إحصائية. وتشمل إحدى فئات هذه التقنيات حساب المسافات بين كل نجم وعدد قليل من النجوم الأقرب إليه. والنجوم التي يكون عدد النجوم القريبة منها أكبر مما هو متوقَّع تكون واقعة في مناطق كثيفة محليًّا؛ أي إنها تشكِّل عناقيد محلية.

بالطبع، يتعلق الأمر بأكثر من ذلك بكثير؛ فسُحُب الغبار بين النجوم ستَحجب رؤية الأشياء البعيدة، وسحب الغبار هذه ليست موزَّعة على نحو موحَّد في الفضاء. وبالمثل، لن تُرى الأجرام الباهتة إلَّا إذا كانت قريبة بما فيه الكفاية من الأرض. والخيط الرفيع من المجرات الذي ترى نهايته من الأرض يمكن أن يبدو كعنقود كثيف، وهكذا. وينبغي تطبيق تصحيحات إحصائية متطورة حتى نتمكن من تمييز الحقيقة الكامنة من التوزيعات الظاهرية للأجرام السماوية.

إن فَهْم بنية الكون يُلقِي الضوء على كيفية تشكُّله، وعلى تطوُّره المستقبلي.

مثال ٤: تصنيع المواد الكيميائية

أشرتُ بالفعل إلى أنه في حين أن الإحصائيين ربما يكونون قادرين على القيام بأمور مذهلة، فإنهم لا يمكن أن يحققوا معجزات؛ وبالتحديد، سوف تتحدد جودة استنتاجاتهم دومًا بجودة البيانات. في ضوء هذا الأمر، ليس من المستغرب وجود تخصصات فرعية مهمة في الإحصاء معنية بأفضل السُّبُل لجمع البيانات، وتُناقَش هذه التخصصات الفرعية في الفصل الثالث. يتمثل أحد هذه التخصصات الفرعية في «التصميم التجريبي»، وتُستخدم تقنيات التصميم التجريبي في الحالات التي من الممكن فيها التحكم أو التلاعب في بعض «المتغيرات» الخاضعة للدراسة. وتُمكِّننا أدوات التصميم التجريبي من استخراج أقصى قدر من المعلومات بالنسبة لأي استخدام معين للموارد؛ فعلى سبيل المثال، في إنتاج بوليمر كيميائي معين ربما نكون قادرين على ضبط درجة الحرارة والضغط ووقت التفاعل الكيميائي بأي قِيَم نريدها. والقِيَم المختلفة لهذه المتغيرات الثلاثة ستؤدي إلى اختلافات في جودة المنتج النهائي. والسؤال هو: ما هي أفضل مجموعة من القيم؟

مبدئيًّا، هذا سؤال يَسهُل الإجابة عنه؛ فنصنع ببساطة العديد من كميات البوليمر، لكلٍّ منها قِيَم مختلفة من المتغيرات الثلاثة. وهذا يسمح لنا بتقدير «استجابة السطح»،

والتي تبيِّن جودة البوليمر عند كل مجموعة من قِيَم المتغيرات الثلاثة، ويمكننا بعد ذلك اختيار القِيَم الثلاث المحددة التي تزيد الجودة إلى الحد الأقصى.

ولكن ماذا لو كانت عملية التصنيع من النوع الذي يستغرق عدة أيام لصنع كل كمية؟ إن صنع العديد من هذه الكميات لمجرد التوصل إلى أفضل طريقة للقيام بذلك ربما يكون أمرًا صعب التنفيذ؛ فصنع مائة كمية، يستغرق صنع كلٍّ منها ثلاثة أيام، سيستغرق الجزء الأكبر من عام كامل. لحسن الحظ، التجارب المصمَّمة بذكاء تسمح لنا باستخراج المعلومات نفسها من مجموعات مختارة بعناية من القِيَم عددها أقلُّ بكثير. وفي بعض الأحيان يمكن لنسبة ضئيلة من الكميات أن تمنحنا معلومات كافية لتحديد أفضل مجموعة من القِيَم، شريطة أن تُختار تلك الكميات على نحو صحيح.

مثال ٥: رضا العملاء

إن إدارة أي مؤسسة للبيع بالتجزئة على نحو فعَّال، بحيث تحقق ربحًا وتنمو مع مرور الوقت، تتطلب إيلاء اهتمام دقيق للعملاء، ومنحَهم المنتج أو الخدمة التي يريدونها. والفشل في القيام بذلك يعني أنهم سيتوجهون إلى مُنافِس يقدِّم ما هو مطلوب. بيت القصيد هنا هو أن الفشل سوف يتضح من انخفاض الإيرادات. ويمكننا محاولة تجنُّب ذلك من خلال جمع بيانات حول مشاعر العملاء قبل أن يبدءوا التصويت بأموالهم. ويمكننا تنفيذ دراسات مسحية لرضا العملاء، سائلين العملاء ما إذا كانوا سعداء بالمنتج أو الخدمة أم لا، وعن الطرق التي يمكن من خلالها تحسين ذلك.

للوهلة الأولى قد يبدو أنه من الضروري منح الاستبيانات لجميع العملاء من أجل الحصول على نتائج موثوقة تعكس سلوك قاعدة العملاء بأكملها، لكن من الواضح أن هذه عملية مكلِّفة وتستغرق وقتًا طويلًا. ومع ذلك، توجد — لحسن الحظ — أساليب إحصائية تُمكِّن من الحصول على نتائج دقيقة بما فيه الكفاية من عيِّنة من العملاء فحسب. وفي الواقع، يمكن أن تكون النتائج أحيانًا أكثر دقة من إشراك جميع العملاء. ولا حاجة بنا لقول إنه يلزم وجود عناية كبيرة في هذه العملية؛ فمن الضروري أن نكون حَذِرين من بناء استنتاجات على عينة مشوهة؛ فربما ستكون النتائج غير مُجدِية في وصف كيفية تصرف العملاء عمومًا إذا أُجريت المقابلات مع أولئك الذين ينفقون مبالغ كبيرة من المال فحسب. ومرة أخرى، طُورت الأساليب الإحصائية التي تمكِّننا من تجنُّب مثل هذه الأخطاء؛ ومن ثم استخلاص استنتاجات صحيحة.

مثال ٦: كشف الاحتيال ببطاقات الائتمان

ليست كل معاملات بطاقات الائتمان شرعية. والمعاملات الاحتيالية تكلف البنك أموالًا، وكذلك تكلف عملاء البنك أموالًا؛ ومن ثم فإن كشف الاحتيال ومَنْعه أمرٌ مُهِمٌّ للغاية. ربما مر العديد من قُرَّاء هذا الكتاب بتجربةِ تلقِّي اتصال هاتفي من المصرف للتأكد من أنهم قاموا ببعض المعاملات. تستند هذه المكالمات الهاتفية على توقعات تقدمها نماذج إحصائية تحدد مدى شرعية تصرفات العملاء. والخروج عن السلوك الذي تتنبَّأ به هذه النماذج يُشِير إلى أن شيئًا مريبًا يجري ويستحق التحقق منه.

توجد أنواع عديدة من النماذج، يعتمد بعضها ببساطة على أنماط السلوك المثيرة للشكوك في جوهرها؛ مثل الاستخدام المتزامن لبطاقة واحدة في مكانين بعيدَيْن جغرافيًّا. ويستند البعض الآخَر على نماذج أكثر تفصيلًا لأنواع المعاملات التي يقوم بها الشخص عادة، ومتى يميل إلى القيام بها، وكمية المال المستخدم، وفي أي أنواع المنافذ، ولأي أنواع المنتجات، وما شابه ذلك.

بطبيعة الحال، لا يوجد نموذج تنبُّئي كامل؛ فغالبًا ما تتنوع أنماط معاملات بطاقة الائتمان؛ حيث إن الناس قد يشترون فجأة منتجات لم يَشتَرُوها من قبلُ. علاوة على ذلك، نسبة ضئيلة فحسب من المعاملات تكون احتيالية؛ ربما حوالي واحد في الألف. وهذا يجعل كشف الاحتيال شديد الصعوبة.

إن كشف الاحتيال ومنعه معركة مستمرة؛ فعند سدِّ أَحَدِ سُبُل الاحتيال، فإن المحتالين لا يميلون إلى التخلِّي عن مسارهم الذي اختاروه والحصول على وظيفة مشروعة، بل يتجهون إلى أساليب أخرى للاحتيال؛ ومن ثم فإن ذلك يتطلب تطوير المزيد من النماذج الإحصائية.

مثال ٧: التضخم

إننا جميعًا نألف فكرة أن الأشياء تزداد غلاءً بمرور الوقت. ولكن كيف يمكننا مقارنة تكاليف المعيشة اليوم بتكاليف المعيشة أمس؟ للقيام بذلك، نحتاج إلى مقارنة الأشياء نفسها التي اشتريناها في اليومين. لكن للأسف، توجد تعقيدات؛ فالمحلات التجارية المختلفة تحدد أسعارًا مختلفة للأشياء نفسها، والأشخاص المختلفون يشترون أشياء مختلفة، ويغير الأشخاص أنفسهم أنماط شرائهم، وتَظهَر منتجات جديدة في السوق وتختفي

منتجات قديمة، وما شابه ذلك. كيف نضع مثل هذه التغييرات في الاعتبار عند تحديدِ ما إذا كانت الحياة أكثر تكلفة هذه الأيام أم لا؟

أنشأ الإحصائيون والاقتصاديون مؤشرات مثل «مؤشر أسعار التجزئة» و«مؤشر أسعار المستهلك» لقياس تكاليف المعيشة. وتستند هذه المؤشرات إلى «سلَّة» افتراضية للسلع (مئات منها) التي يشتريها الناس، إضافة إلى دراسات استقصائية لاكتشاف الأسعار التي يُباع بها كل عنصر في السلة. وتُستخدم نماذج إحصائية متطورة لجمع أسعار العناصر المختلفة لتقدم رقمًا إجماليًّا واحدًا يمكن مقارنته على مدار الزمن. وبالإضافة إلى كونها مؤشرًا على التضخم، تستخدم هذه المؤشرات أيضًا لضبط حدود الإعفاء الضريبي والرواتب المرتبطة بالمؤشر والمعاشات التقاعدية، وما إلى ذلك.

خاتمة

رغم أن هذا قد لا يبدو واضحًا دائمًا للعين غير الخبيرة، فإن علم الإحصاء والأساليب الإحصائية يَكْمُنان في قلب الاكتشاف العلمي، والعمليات التجارية والحكومية، والسياسة الاجتماعية، والتصنيع، والطب، ومعظم جوانب النشاط الإنساني الأخرى. علاوة على ذلك، كلما تقدم العالَم، زادت أهمية هذا الدور أكثر وأكثر؛ على سبيل المثال، منذ وقت طويل وتطويرُ أدوية جديدة يشترط، قانونًا، مشاركةَ الإحصائيين، وشيء من هذا القبيل يحدث الآن في الصناعة المصرفية؛ حيث إن الاتفاقات الدولية الجديدة تتطلب وضع نماذج إحصائية للمخاطر. ونظرًا لهذا الدور المحوري، من المهم بوضوح أن يكون أي مُواطن مستنير على علم بالمبادئ الإحصائية الأساسية.

يمكِّننا علم الإحصاء الحديث، الذي يستخدم البرمجيات المتطورة لدراسة البيانات، من القيام برحلات استكشاف مشابهة لتلك التي قام بها المستكشفون قبل القرن العشرين؛ إذ استقصَوْا ودرسوا عوالم جديدة ومثيرة. وهذا الإدراك — أن علم الإحصاء الحقيقي يتمحور حول استكشاف المجهول، ولا يتمحور حول عمليات حسابية مُمِلَّة — أساسيٌّ في تقدير قيمة هذا العلم الحديث.

الفصل الثاني

تعريفات بسيطة

البيانات أدلة الطبيعة.

مقدمة

أهدف في هذا الفصل إلى تقديم بعض المفاهيم والأدوات الأساسية التي تشكل أساس علم الإحصاء، والتي تمكِّنه من لعب أدوار كثيرة.

أشرتُ في الفصل الأول إلى أنَّ علم الإحصاء الحديث عانَى من كثير من المفاهيم الخاطئة وسوء الفهم. ومع ذلك، يروَّج سوء فهم آخَر في كثير من الأحيان (ربما عن غير قصد) عن طريق الكتب التي تشرح الأساليب الإحصائية للخبراء في تخصصات أخرى؛ وهو أن الإحصاء عبارة عن حقيبة من الأدوات، ويتمثل دور الإحصائي أو مستخدِم الإحصاء في اختيار أداة واحدة تتناسب مع مسألته، ثم تطبيقها.

تتمثل مشكلة هذه النظرة للإحصاء في أنها تعطي انطباعًا بأن مجال الإحصاء ببساطةٍ عبارة عن مجموعة من الطرق المنفصلة لمعالجة الأرقام؛ فهي تفشل في نقل حقيقة أن الإحصاءَ كلٌّ متصلٌ، مبني على مبادئ فلسفية عميقة، بحيث تكون أدوات تحليل البيانات مرتبطة ومتصلة؛ فبعضها قد يبدو شاملًا مقارنة بغيره، وربما يبدو البعض الآخر مختلفًا ببساطة لأنه يتعامل مع أنواع مختلفة من البيانات، على الرغم من أن هذه الأدوات تبحث عن النوع نفسه من البِنَى، وما إلى ذلك. وأظن أن انطباع مجموعة الطرق المعزولة هذا ربما يكون سببًا آخر يدفع المستجدِّين في مجال الإحصاء إلى الاعتقاد بأن هذا المجال مملٌّ نوعًا ما وصعب التعلم (بصرف النظر عن أي خوف من الأرقام قد يكون لديهم)؛ فتعلم مجموعة من الطرق المنفصلة التي تبدو شديدة التباين أصعب بكثير من تعلم هذه الطرق من خلال اشتقاقها من المبادئ الأساسية نفسها.

الأمر يشبه في صعوبته تعلُّم مجموعة عشوائية من الكلمات غير المرتبطة، مقارنة بتعلم كلماتِ جملةٍ ذات معنًى. ولقد سعيتُ — في هذا الفصل وعلى مدار الكتاب — للتعبير عن العلاقات بين الأفكار الإحصائية، من أجل إيضاح أن مجال الإحصاء في الحقيقة وحدة متكاملة مترابطة.

(١) البيانات مرة أخرى

أيًّا كان ما يفعله علم الإحصاء، وبغضِّ النظر عن تفاصيل التعريف الذي نعتمده له، فإن علم الإحصاء يبدأ بالبيانات. تصف البيانات الكون الذي نرغب في دراسته، وأستخدم كلمة «الكون» هنا بمعنًى عام واسع؛ فيمكن أن يكون العالَم المادي الذي يدور حولنا، ويمكن أيضًا أن يكون عالَم معاملات بطاقات الائتمان، أو عالم تجارب المصفوفات الدقيقة في علم الوراثة، أو عالم المدارس والتدريس وأداء الامتحانات، أو عالم التجارة بين البلدان، أو عالم كيفية تصرف الأشخاص عند التعرض للإعلانات المختلفة، أو عالم الجسيمات دون الذرية، وما شابه ذلك. لا توجد نهاية للعوالم التي يمكن دراستها؛ ومن ثم لا نهاية للعوالم التي تمثلها البيانات.

بطبيعة الحال، لا يمكن لمجموعة محدودة من البيانات أن تُخبِرنا عن كل التعقيدات اللانهائية للعالَم الحقيقي، تمامًا كما لا يوجد وصفٌ لفظي — حتى إنْ كَتَبَه أفصح المؤلِّفين — يمكن أن ينقل كل شيء عن كل جانب من جوانب العالم من حولنا؛ وهذا يعني أننا يجب أن نكون واعِين للغاية بأي مَوَاطن ضعف أو ثغرات في البيانات لدينا، ويعني أنه عند جمع البيانات، نكون بحاجة لإيلاء عناية خاصة للتأكد من أنها تغطِّي بالفعل الجوانب التي نهتم بها، أو التي نرغب في استخلاص نتائج حولها. توجد أيضًا طريقة أكثر إيجابية للنظر إلى هذا الأمر؛ وهي أنه عن طريق جمع مجموعة محدودة من الجوانب الوصفية فحسب، فإننا نُضطر لإقصاء العناصر غير ذات الصلة؛ فعند دراسة سلامة تصميمات السيارات المختلفة، ربما نقرر عدم تسجيل لون القماش الذي يكسو المقاعد.

من الملائم عمومًا النظر للبيانات على أن لها جانبَيْن؛ يتعلَّق أحدهما بالكائنات التي نرغب في دراستها، ويتعلق الجانب الآخر بخصائص هذه الكائنات التي نرغب في دراستها؛ على سبيل المثال، ربما تتمثل هذه الكائنات في أطفال المدرسة وتتمثل خصائصهم في درجاتهم في الاختبار، أو ربما تتمثل الكائنات في الأطفال، ولكننا ندرس

نظامهم الغذائي ونموِّهم البدني، وفي هذه الحالة ربما تتمثل الخصائص في طول الأطفال ووزنهم، أو ربما تكون هذه الكائنات موادَّ مادية، أما الخصائص ذات الأهمية فهي سماتها الكهربائية والمغناطيسية. من الشائع في مجال الإحصاء تسمية هذه الخصائص «متغيرات»، بحيث يمتلك كل كائن منها «قيمة» للمتغير (درجة الطفل في اختبار الإملاء تمثل قيمة متغير الاختبار، وكمية التوصيل الكهربي للمادة تمثل قيمة متغير القدرة على توصيل التيار، وما إلى ذلك). وفي مجالات تحليل البيانات الأخرى، تُستَخدَم كلمات بديلة في بعض الأحيان (مثل «ميزة» أو «سمة» أو «خاصية»)، ولكن عند مناقشة الجوانب التقنية، سألتزم عادة بكلمة «متغير».

في الواقع، في أي دراسة، ربما نكون مهتمين بأنواع متعددة من الكائنات. فربما لا نرغب في الفهم وتقديم النتائج عن أطفال المدارس فحسب، ولكن أيضًا عن المدارس نفسها وربما عن المعلمين وأساليب التدريس والأنواع المختلفة لهياكل الإدارة المدرسية، كل ذلك في دراسة واحدة. علاوة على ذلك، عادة لن نكون مهتمين بسمة واحدة للكائنات التي تخضع للدراسة، وإنما بالعلاقات بين السمات، وربما بالفعل بالعلاقات بين سمات الكائنات من الأنواع المختلفة وعلى المستويات المختلفة. وكما هو متوقَّع، نجد أن الأمور غالبًا ما تكون معقَّدة للغاية؛ نظرًا لتعقيد الموضوعات التي ندرسها.

يقاوم كثير من الناس فكرة أنه يمكن للبيانات الرقمية أن تنقل جمال العالم الحقيقي؛ فيشعرون بأن تحويل الأشياء إلى أرقام يُزِيل بطريقة أو بأخرى عنها سحرها. في الواقع، هم مخطئون حتى النخاع؛ فالأرقام لديها القدرة على السماح لنا بإدراك هذا الجمال — هذا السحر — على نحو أكثر وضوحًا وأكثر عمقًا، وتقديره حقَّ قدْره. وباعتراف الجميع، ربما يُزال «الغموض» عن طريق وصف الأشياء بصورة رقمية؛ فإذا قلتُ إنه يوجد أربعة أشخاص في الغرفة، فإنك تعرف بالضبط ما أَعنِيه، في حين أنني إذا قلتُ إن شخصًا ما جذَّاب، ربما لا تكون متأكدًا تمامًا ممَّا أعنِيه. وربما تختلف حتى مع وجهة نظري في أن ثمة شخصًا جذَّابًا في الغرفة، ولكن من غير المرجَّح أن تختلف مع وجهة نظري بأن هناك أربعة أشخاص في الغرفة (باستثناء أخطاء العدِّ بطبيعة الحال، ولكن هذا أمر مختلف). والأرقام مفهومة على نحو عالمي، بغض النظر عن الجنسية أو الدين أو الجنس أو العمر أو أي سِمَة بشرية أخرى. ويمكن أن تكون إزالة الغموض — ومعها إزالة خطر سوء الفهم — مفيدة عندما نحاول أن نفهم شيئًا؛ عندما نحاول فهمه تمامًا.

ويرتبط افتقاد الغموض هذا في تفسير الأرقام ارتباطًا وثيقًا بحقيقة أن «الأرقام تمتلك سمة واحدة فقط»؛ ونعني بهذا قيمتها أو حجمها. فعلى النقيض مما قد يدفعنا العرافون إلى الإيمان به، فإن الأرقام ليست جالِبة للحظ الجيد أو السيئ؛ تمامًا كما أن الأرقام لا تمتلك لونًا أو نكهة أو رائحة، فليس لديها سمات غير قيمتها الرقمية الذاتية. (لا يمكن إنكار أن بعض الأشخاص يمتلكون «الحس المرافق»، والذي فيه يربطون لونًا معينًا أو إحساسًا بأرقام معينة. ومع ذلك، فإن الأحاسيس المرتبطة تتباين باختلاف الأشخاص، ولا يمكن اعتبارها سمات خاصة بالأرقام نفسها.)

تقدِّم البيانات الرقمية لنا صلة مباشرة وفورية بالظواهر التي ندرسها أكثر مما تقدِّمه الكلمات؛ لأن البيانات الرقمية تَنتج عادة عن طريق أدوات قياس تتصل اتصالًا مباشرًا بتلك الظواهر بدرجة أكبر من اتصالها بالكلمات؛ فالأرقام تأتي مباشرة من الأشياء التي تجري دراستها، في حين أن الكلمات تخضع للترشيح عن طريق العقل البشري. بطبيعة الحال، فإن الأشياء تكون أكثر تعقيدًا إذا تمَّتْ إجراءات جمع البيانات بواسطة الكلمات (كما هي الحال إذا جُمِعت البيانات عن طريق الاستبيانات)، ولكن لا يزال المبدأ صالحًا. وبينما قد لا تكون أدوات القياس مثالية، فإن البيانات تكون تمثيلًا حقيقيًّا لنتائج تطبيق تلك الأدوات على الظاهرة قيد الدراسة. وأحيانًا ألخص ذلك من خلال التعليق الموجود في بداية هذا الفصل: «البيانات هي أدلة الطبيعة، التي تُرى من خلال عدسة أداة القياس.»

وفوق كل هذا، للأرقام نتائج عملية من حيث التقدم المجتمعي؛ فقدرة العالَم المتحضِّر على معالجة تمثيلات الواقع التي تقدِّمها الأرقام هي التي أدَّت إلى مثل هذا التقدم المادي المذهل في القرون القليلة الماضية.

على الرغم من أن الأرقام لها سمة واحدة فقط — قيمتها الرقمية — فربما نختار استخدام تلك السمة بطرق مختلفة؛ على سبيل المثال، عند اتخاذ قرار بشأن جدارة الطلاب في الصف الدراسي، ربما نصنِّفهم وفقًا لدرجات الامتحان؛ أي إننا ربما لا نهتمُّ إلَّا بما إذا كانت نتيجةٌ ما أعلى من أخرى، ولا نهتمُّ بالفارق العددي الدقيق. وعندما نهتم فقط «بترتيب» القِيَم بهذه الطريقة نقول إننا نعالج البيانات بوضعها على مقياس «ترتيبي». من ناحية أخرى، عندما يقيس المزارع كمية الذرة التي أنتجها، فلا يريد ببساطة معرفةَ ما إذا كان قد أنتج أكثر مما أنتج في العام الماضي أم لا، كما أنه يريد أيضًا أن يعرف مقدارَ ما أنتجه؛ أي الوزن الفعلي؛ فعلى أي حال، سوف تُباع الذرة في

السوق على هذا الأساس. في هذه الحالة، يُقارِن المزارع فعليًّا وزن الذرة التي أنتجها بوزن معياري مثل الطن، حتى يستطيع معرفة كم طنًّا من الذرة أنتجه. يتضمن ذلك احتساب نسبة وزن الذرة التي أنتجها المزارع لوزن الطن الواحد من الذرة؛ لهذا السبب، عندما نستخدم القِيَم على هذا النحو، فإننا نقول إننا نعالج البيانات بوضعها على مقياس «نسبي». لاحِظْ أنه في هذه الحالة يمكننا اختيار تغيير وحدة القياس الأساسية؛ إذ يمكننا حساب الوزن بالرطل أو الكيلوجرام بدلًا من الطن. وما دمنا نشير إلى الوحدة التي استخدمناها، فإنه من السهل على أي شخص آخر إعادة تحويلها مرة أخرى، أو تحويلها إلى أي وحدة يستخدمها عادة.

في حالة أخرى، ربما نرغب في معرفة عدد المرضى الذين عانَوْا من أثر جانبي معين لدواءٍ ما. وإذا كان العدد كبيرًا بما فيه الكفاية فإننا قد نرغب في سَحْب الدواء من السوق على أساس أنه ينطوي على مخاطرة كبيرة للغاية. في هذه الحالة، فإننا ببساطة نُحصي الوحدات المنفصلة الواضحة المعالِم (المرضى). لن تكون إعادة القياس عن طريق تغيير الوحدات ذات مغزًى (فلن نفكِّر في إحصاء عدد «نصف المرضى»!) لذلك نقول إننا نُعالِج البيانات بوضعها في المقياس «المطلق».

(٢) الملخَّصات الإحصائية البسيطة

في حين أن الأرقام البسيطة تشكل «عناصر» البيانات، فإنه من أجل أن تكون مفيدة، فإننا نحتاج إلى أنْ ننظر في العلاقات بينها، وربما نَجمَع بينها بطريقةٍ ما، وهنا يأتي دَوْر الإحصاء. سوف تستكشف الفصول اللاحقة طرقًا أكثر تعقيدًا لمقارنة الأرقام والجمع بينها، ولكن سيكون هذا الفصل بمنزلة مقدمة للأفكار. سنُلقِي هنا نظرة على بعض أكثر الطرق مباشرة؛ فلن نستكشف العلاقات بين المتغيرات المختلفة في هذا الفصل، ولكن ببساطة سنرى المعلومات والرُّؤَى التي يمكن استخلاصها من العلاقات بين القِيَم المَقيسة وفْق المتغير نفسه؛ على سبيل المثال، ربما نكون قد سجَّلنا أعمار المتقدِّمين للحصول على منصب في الجامعة، أو درجة سطوع النجوم في عنقود مَجَرِّيٍّ ما، أو النفقات الشهرية للأُسَر في مدينةٍ ما، أو أوزان أبقار في قطيع في وقت إرسالها إلى السوق، وما إلى ذلك. وفي كل حالة، تُسجَّل قيمة رقمية واحدة لكل «كائن» في مجموعة الكائنات.

عندما تؤخذ معًا، يُقال إن القِيَم الفردية في المجموعة تشكِّل «توزيعًا» للقِيَم. وتُعَدُّ الملخصات الإحصائية سبلًا لتمييز هذا التوزيع؛ أي قول ما إذا كانت القِيَم متشابِهة

جدًّا، وما إذا كانت توجد بعض القِيَم الكبيرة أو الصغيرة على نحو استثنائي، وتحديد القيمة «النموذجية» ... إلخ.

(٢-١) القِيَم المتوسطة

يتمثل أبسط أنواع التوصيفات — أو الملخصات الإحصائية — لمجموعة من الأرقام في «القيمة المتوسطة». والقيمة المتوسطة هي قيمة تمثيلية؛ وهي قيمة قريبة، بمعنًى ما، لأرقام المجموعة. والحاجة إلى شيء من هذا القَبِيل تكون أكثر وضوحًا عندما تكون مجموعة الأرقام كبيرة؛ على سبيل المثال، لنفترض أن لدينا جدولًا يسجِّل أعمار كل الأشخاص في مدينة كبيرة؛ ربما يبلغ عددهم مليون نسمة. من أجْل الأغراض الإدارية والتجارية سيكون من المفيد على نحو واضح معرفة متوسط عمر السكان؛ فسوف توجد حاجة لخدمات مختلفة للغاية، وتنشأ فرص مبيعات إذا كان متوسط العمر ١٦ عامًا بدلًا من ٦٠. وبإمكاننا أن نحاول الحصول على فكرة عن الحجم العام للأرقام في الجدول — الأعمار — من خلال النظر إلى كل القِيَم. لكن من الواضح أن هذا سيكون أمرًا عسيرًا. في الواقع، إذا كان النظر إلى كل رقم يستغرق ثانية واحدة فقط، فإن الأمر سيستغرق أكثر من ٢٧٠ ساعة للنظر إلى جدول مكوَّن من مليون رقم، كل هذا مع تجاهل العمل الفعلي المتمثل في محاولةِ تذكُّرِها ومقارنتها. ولكنْ يمكننا استخدام جهاز الكمبيوتر الخاص بنا لمساعدتنا.

أولًا: نحن بحاجة إلى أن نكون واضِحِين حيالَ ما نَعنِيه بكلمة «قيمة متوسطة» بالضبط، لأن الكلمة لها عِدَّة معانٍ. ربما النوع الأكثر استخدامًا من القيمة المتوسطة هو «المتوسط الحسابي»، أو «الوسط الحسابي». فإذا استخدم الشخص كلمة «المتوسط» دون أن يوضح تفسيرها، فإنه ربما حينها يكون قاصدًا «المتوسط الحسابي».

وقبل أن أوضح كيفية حساب المتوسط الحسابي، تَخيَّلْ جدولًا آخر يحتوي مليون رقم. لنفترضْ في هذا الجدول الثاني أن جميع الأرقام متطابقة بعضها مع بعض؛ أي لنفترضْ أنها جميعًا لها القيمة نفسها. والآن اجمع جميع الأرقام في الجدول الأول لإيجاد مجموعها الكلي (هذا لا يستغرق سوى جزء من الثانية باستخدام جهاز كمبيوتر). اجمع جميع الأرقام في الجدول الثاني لإيجاد مجموعها الكلي. إذا كان مجموعا أرقام الجدولين بالقيمة نفسها، فإن الرقم الذي تكرر مليون مرة في الجدول الثاني يمثل قيمة جوهرية نوعًا ما بالنسبة للأرقام في الجدول الأول. هذا الرقم المفرد، والذي جمعت منه مليون

نسخة لتصل إلى المجموع نفسه كما في الجدول الأول، يسمَّى المتوسط الحسابي (للأرقام في الجدول الأول).

في الواقع، أسهل السُّبُل لحساب المتوسط الحسابي هي من خلال قسمة مجموع الأرقام المليون في الجدول الأول على مليون. وعمومًا، يتم إيجاد المتوسط الحسابي لمجموعة من الأرقام بجمع جميع الأرقام وقسمة المجموع على عددها. إليك مثالًا آخَر: في اختبارٍ ما، كانت النسبة المئوية لنتائج خمسة طلاب في الصف هي: ٧٨، ٦٣، ٥٣، ٩١، ٥٥. يبلغ مجموع هذه الأرقام: ٧٨ + ٦٣ + ٥٣ + ٩١ + ٥٥ = ٣٤٠. ويأتي المتوسط الحسابي ببساطة عن طريق قسمة ٣٤٠ على ٥؛ وهو ٦٨. وكنا سنحصل على المجموع نفسه (٣٤٠) إذا حصل جميع الطلاب الخمسة على القيمة المتوسطة ٦٨.

يمتلك المتوسط الحسابي العديد من الخصائص الجذابة؛ فدائمًا ما يأخذ قيمة بين أكبر القِيَم وأصغرها في مجموعة الأرقام. علاوة على ذلك، فإنه يوازن بين الأرقام في المجموعة؛ بمعنى أن مجموع الفروق بين المتوسط الحسابي والقِيَم الأكبر منه يساوي بالضبط مجموع الفروق بين المتوسط الحسابي والقِيَم الأصغر منه. وبهذا المعنى، هو قيمة «مركزية». والأشخاص الذين يملكون تفكيرًا ميكانيكيًّا قد يرغبون في تصور مجموعة من الأثقال زِنَة الواحد منها كيلوجرام واحد موضوعة في مواقع مختلفة على طول لوح خشبي (عديم الوزن). ومسافات الأوزان من أحد طرفي اللوح تمثِّل القِيَمَ في مجموعة الأرقام. والمتوسط هو المسافة التي تفصل الطرف عن محور ارتكاز يتوازن فيه لوح الخشب تمامًا.

المتوسط الحسابي هو «إحصائية»، وهو يلخص مجموعة كاملة من القيم في مجموعتنا في صورة قيمة واحدة. يتبع ذلك أنه يهمل أيضًا معلومات؛ فيَجِب ألَّا نتوقَّع أن نُمثِّل مليون رقم مختلف (أو خمسة، أو أيًّا كان عددها) عن طريق رقم واحد دون التضحية بشيءٍ ما، وسنعمل على استكشاف هذه التضحية في وقت لاحق. ولكنْ نظرًا لأنه قيمة مركزية بالمعنى المُبَيَّن أعلاه، فإنه يمكن أن يكون ملخصًا مفيدًا؛ فيمكننا مقارنة متوسطات حجم الفصل في المدارس المختلفة، أو متوسط درجة اختبار طلاب مختلفين، أو متوسط الوقت الذي يستغرقه مختلف الناس للوصول إلى العمل، أو متوسط درجة الحرارة اليومية في سنوات مختلفة، وما إلى ذلك.

المتوسط الحسابي إحصائية مهمة؛ فهو ملخص لمجموعة من الأرقام. وثمة ملخص آخر مهم هو «الوسيط». كان المتوسط هو القيمة المحورية؛ نوعًا من النقطة المركزية

الموازنة لمجموع الفروق بينه وبين الأرقام في المجموعة. أما الوسيط فيوازن المجموعة بطريقة أخرى؛ فهو القيمة التي يكون نصف الأرقام في مجموعة البيانات أكبر منها والنصف الآخر أصغر منها. وبالعودة إلى الصف المكوَّن من خمسة طلاب المذكور أعلاه، فإن نتائجهم بالترتيب من الأصغر إلى الأكبر هي: ٥٣، ٥٥، ٦٣، ٧٨، ٩١. والنتيجة الوسطى هنا هي ٦٣، لذلك هذا هو الوسيط.

من الواضح أنه ستظهر بعض التعقيدات إذا وجدت قيم متساوية في مجموعة البيانات (لنفترض على سبيل المثال أنها تتكون من ٩٩ نسخة من القيمة ٠ ونسخة واحدة من القيمة ١)، ولكن يمكن التغلب على ذلك. على أي حال، مرة أخرى الوسيط هو قيمة تمثيلية بمعنًى ما، وإن كان يختلف عن المتوسط. وبسبب هذا الاختلاف، لنا أن نتوقع أنه سيأخذ قيمة مختلفة عن المتوسط. من الواضح أن الوسيط أسهل في الحساب من المتوسط. فليس علينا جمع أي قيم للوصول إليه، فضلًا عن القسمة على عدد القيم في المجموعة؛ كل ما عليك القيام به هو ترتيب الأرقام، وتحديد موقع الرقم الموجود في الوسط. ولكن في الواقع هذه الميزة الحسابية أساسًا غيرُ ذات صلة بعصر الكمبيوتر؛ ففي التحليلات الإحصائية الحقيقية يقوم الكمبيوتر بعمليات المعالجة الحسابية المملة.

بوجود هذين الملخصين الإحصائيين، وكلاهما يقدِّم قِيَمًا تمثيلية، كيف لنا أن نحدد أيهما سنستخدم في أي موقف معين؟ بما أنهما يُعرَّفان على نحو مختلف — يجمعان القيم الرقمية على نحو مختلف — فمن المرجَّح أن ينتجا قيمًا مختلفة؛ ولذلك ربما تكون أي استنتاجات تستند إليهما مختلفة للغاية. والجواب الكامل لمسألة أيهما تختار سوف يدخلنا في أمور فنية تتجاوز مستوى هذا الكتاب، ولكن الجواب القصير هو أن الاختيار سيعتمد على التفاصيل الدقيقة للسؤال الذي يرغب المرء في الإجابة عنه.

إليك مثالًا: لنفترض أن شركة صغيرة لديها خَمس مجموعات من الموظفين، لكلٍّ منها درجة ومرتب مختلفان؛ وهي على الترتيب: ١٠٠٠٠ دولار، ١٠٠٠١ دولار، ١٠٠٠٢ دولار، ١٠٠٠٣ دولارات، ٩٩٩٩٩ دولارًا. متوسط هذه القِيَم هو ٢٨٠٠١ دولار، في حين أن الوسيط هو ١٠٠٠٢ دولار. والآن لنفترض أن الشركة تعتزم توظيف خمسة موظفين جُدُد؛ واحد لكل درجة. ربما يُشِير صاحب العمل إلى أنه في هذه الحالة، سيُضطر «في المتوسط» لدفع راتب إجمالي للقادمين الجُدُد الخمسة كلهم يبلغ ٢٨٠٠١ دولار؛ ومن ثم يكون هذا هو متوسط الراتب الذي يذكره في الإعلان. لكن ربما يشعر الموظفون أن هذا تحايل؛ لأن عدد الموظفين الذين سيُدفَع لهم أقل من ١٠٠٠٢ دولارات سيساوي عدد

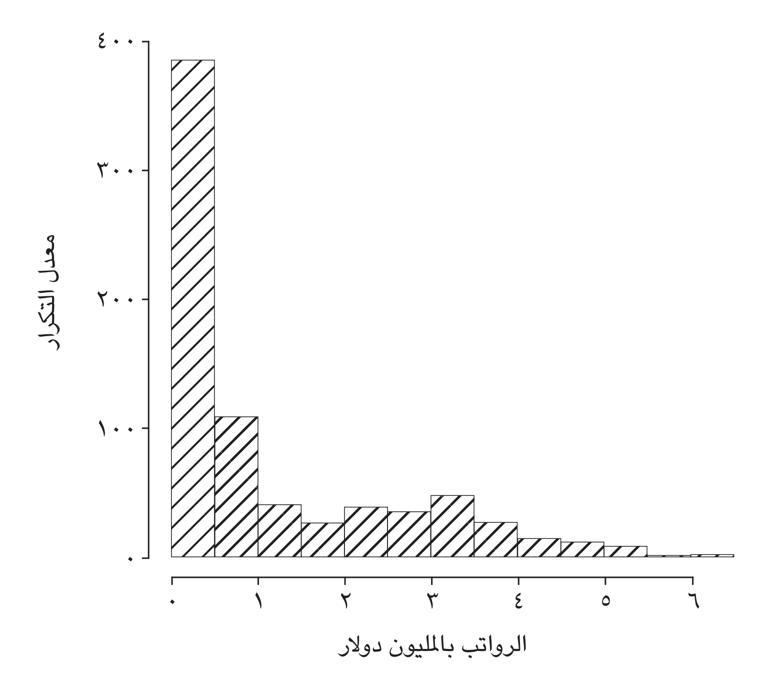
الموظفين الذين سيُدفَع لهم مبلغ أكثر من ١٠٠٠٢ دولارات. وربما يشعرون أنه من الأكثر صدقًا وضع هذا الرقم في الإعلان. أحيانًا يتطلب تحديد أي المقياسين هو المناسب تفكيرًا متأنِّيًا. (وفي حال كنتَ تعتقد أن هذه الحجة مبتدَعة، يبيِّن شكل ٢-١ توزيع رواتب لاعبي البيسبول الأمريكي قبل الإضراب في عام ١٩٩٤. كان المتوسط الحسابي ١٫٢ مليون دولار، ولكن كان الوسيط ٠٫٥ مليون دولار.)

يوضح هذا المثال أيضًا التأثير النسبي للقِيَم المتطرِّفة على المتوسط والوسيط. في مثال المرتبات أعلاه، يساوي المتوسط ما يقرب من ثلاثة أضعاف الوسيط. ولكنْ لنفترض أن أكبر قيمة كانت ١٠٠٠٤ دولارات بدلًا من ٩٩٩٩٩ دولارًا، حينها سيظل الوسيط ١٠٠٠٢ دولارات (نِصْف القِيَم أعلاه ونصفها أدناه)، إلَّا أن المتوسط سيتقلَّص إلى ١٠٠٠٢ دولارات. إن حجم قيمة واحدة فقط يمكن أن يكون له تأثير كبير على المتوسط، ولكنه لا يؤثر على الوسيط. وحساسية المتوسط تلك حيال القِيَم المتطرفة هي أحد الأسباب التي تجعل الوسيط أحيانًا مفضلًا في الاختيار عن المتوسط.

ليس المتوسط والوسيط الملخَّصين الوحيدين للقيم التمثيلية؛ فثمة ملخص آخر مهم هو «المنوال»؛ وهو أكثر القيم تكرارًا في العينة؛ على سبيل المثال، لنفترض أنني أُحصِي عدد الأطفال في الأسرة في مجموعة سكانية معينة. ربما أجد أن بعض الأُسَر لديها طفل واحد، وبعضها لديها طفلان، وبعضها ثلاثة، وما إلى ذلك، وربما أجد على وجه الخصوص أن عدد الأُسَر التي لديها طفلان أكبر من أي قيمة أخرى. في هذه الحالة، سيكون منوال عدد الأطفال لكل أسرة هو اثنين.

(٢-٢) التشتت

تقدِّم المتوسطاتُ — على غرار المتوسط الحسابي والوسيط — ملخصاتٍ رقميةً واحدةً لمجموعات من القيم الرقمية، وهي مفيدة لأنها يمكن أن تعطي مؤشرًا عن الحجم العام للقيم الموجودة في البيانات. ولكنْ، كما رأينا في المثال السابق، يمكن للقِيَم التلخيصية الواحدة أن تكون مضلِّلة. وعلى وجه التحديد، قد تنحرف القِيَم التلخيصية الواحدة كثيرًا عن القِيَم الفردية في مجموعة الأرقام. ولتوضيح ذلك، لنفترض أن لدينا مجموعة من مليون رقم وواحد، لها القيم: ٠، ١، ٢، ٣، ٤، …، ١٠٠٠٠٠٠. إن المتوسط والوسيط كليهما لهذه المجموعة من القِيَم يُساوي ٥٠٠٠٠٠. ولكن من الواضح تمامًا أن هذه



شكل ٢-١: توزيع رواتب لاعبي البيسبول الأمريكيين في عام ١٩٩٤. يبين المحور الأفقي الرواتب بالمليون دولار، ويبين المحور الرأسي أعداد اللاعبين في كل نطاق من الرواتب.

القيمة ليست قيمة «تمثيلية» جيدة للمجموعة. فعلى طرفي المجموعة، ثمة قيمة واحدة أكبر بنصف مليون وقيمة واحدة أصغر بنصف مليون من المتوسط (والوسيط).

إن ما نفتقده عندما نعتمد فقط على المتوسط لتلخيص مجموعة من البيانات هو بعض المؤشرات حول مدى انتشار البيانات حول هذا المتوسط؛ هل بعض نقاط البيانات أكبر بكثير من المتوسط؟ هل بعضها أصغر منه بكثير؟ أم إنها متجمعة في تقارب حول المتوسط؟ وعمومًا، ما مدى اختلاف القِيَم في مجموعة البيانات بعضها عن بعض؟ تقدِّم المقاييس الإحصائية للتشتت هذه المعلومات بدقة، وكما هي الحال مع المتوسط، يوجد أكثر من مجرد مقياس واحد.

أبسط مقاييس التشتت هو «المدى»؛ والذي يُعرَّف بأنه الفرق بين أكبر وأصغر القيم في مجموعة البيانات. في مجموعة بياناتنا المكونة من مليون رقم وواحد، المدى هو: ١٠٠٠٠٠٠ − ٠ = ١٠٠٠٠٠٠. وفي مثال الرواتب الخمسة، المدى هو: ٩٩٩٩٩ − ١٠٠٠٠ = ٨٩٩٩٩. ويبيِّن هذان المثالان — اللذان يمتلكان مدًى كبيرًا — أنه يوجد اختلاف كبير عن المتوسط؛ على سبيل المثال، إذا كان الموظفون يتقاضَوْن رواتب تبلغ ٢٧٩٩٩ دولارًا، ٢٨٠٠٠ دولار، ٢٨٠٠١ دولار، ٢٨٠٠٢ دولار، ٢٨٠٠٣ دولارات، فإن المتوسط سيكون أيضًا ٢٨٠٠١ دولار، ولكن سيكون المدى ٤ دولارات فقط. هذا يرسم صورة مختلفة جدًّا تخبرنا أن الموظفين مع هذه الرواتب الجديدة سيتقاضَوْن الأجر نفسه تقريبًا. أما المدى الكبير من المثال السابق — البالغ ٨٩٩٩٩ دولارًا — فيُخبرنا على الفور أنه توجد اختلافات ضخمة.

المدى مقياس ملائم للغاية وله العديد من الخصائص الجذابة كمقياس للتشتت، من أهمها بساطته وإمكانية تفسيره السهلة. ومع ذلك، من الممكن أن نشعر أنه ليس مثاليًّا؛ فهو رغم كل شيء يتجاهل معظم البيانات؛ حيث يعتمد فقط على أكبر القيم وأصغرها. وللتوضيح، تَخيَّلْ مجموعتَين من البيانات تتألف كلٌّ منهما من ألف قيمة. تتضمن إحدى المجموعتَين قيمة واحدة تبلغ ٠، و٩٩٨ قيمة تبلغ ٥٠٠، وقيمة واحدة تبلغ ١٠٠٠. وتتضمن مجموعة البيانات الأخرى ٥٠٠ قيمة تبلغ ٠، و٥٠٠ قيمة تبلغ ١٠٠٠. مدى كلتا مجموعتَي البيانات هو ١٠٠٠ (وبالمصادفة، لكل منهما أيضًا متوسط يبلغ ٥٠٠)، ولكن من الواضح أنهما مختلفتان للغاية في طبيعتهما؛ فبالتركيز فقط على أكبر القيم وأصغرها، فشل المدى في كشف حقيقة أن مجموعة البيانات الأولى تتركز غالبًا بكثافة حول المتوسط.

يمكن التغلب على هذا القصور باستخدام مقياس للتشتت يضع القيم «كلها» في الاعتبار.

إحدى الطرق الشائعة للقيام بذلك هي أن تحسب الفروق بين المتوسط (الحسابي) وكل رقم في مجموعة البيانات، وتقوم بتربيع هذه الفروق؛ ومن ثم تحسب متوسط هذه الاختلافات المربعة. (تربيع الفروق يجعل القِيَم جميعها موجبة. فبخلاف ذلك، سوف تلغي الفروق الموجبة والسالبة بعضها بعضًا عندما نقوم بحساب المتوسط.) وإذا كان المتوسط الناتج عن الفروق المربعة صغيرًا، فإنه يخبرنا في العادة أن الأرقام ليست مختلفة كثيرًا عن متوسطها؛ وهذا يعني أنها ليست مشتتة على نطاق واسع.

يُسمَّى مقياس متوسط الفروق المربعة «تباين» البيانات؛ أو يسمَّى في بعض التخصصات «متوسط مربعات انحرافات القِيَم». وسنوضح الأمر باستخدام درجات الطلاب الخمسة في الاختبار التي كانت ٧٨، ٦٣، ٥٣، ٩١، ٥٥، وكان متوسطها ٦٨. الفارق المربع بين النتيجة الأولى والمتوسط هو (٧٨ – ٦٨)٢ = ١٠٠، وهكذا. ومجموع الفروق المربعة هو ١٠٠ + ٢٥ + ٢٢٥ + ٥٢٩ + ١٦٩ = ١٠٤٨؛ ومن ثم فإن متوسط مربعات انحرافات القيم هو ١٠٤٨ ÷ ٥ = ٢٠٩,٦. وهذا هو التباين.

ينشأ تعقيد طفيف من حقيقة أن التباين ينطوي على قيم مربعة؛ وهذا يعني أن التباين نفسه يقاس بـ «وحدات مربعة». فإذا كنا نقيس إنتاجية المزارع من حيث أطنان الذرة، فإن تباين القيم يُقاس بـ «الطن المربع». ليس تأثير هذا الأمر واضحًا، وبسبب هذه الصعوبة، من الشائع أن نحسب الجذر التربيعي للتباين. وهذا يُعِيد وحدات القياس إلى صورتها الأصلية، ويُنتج مقياسًا للتشتت يُسمَّى «الانحراف المعياري». وفي المثال السابق، يتمثل الانحراف المعياري لدرجات الطلاب في الاختبار في الجذر التربيعي للعدد ٢٠٩,٦، وهو ١٤,٥.

يتغلب الانحراف المعياري على المشكلة التي وجدناها مع المدى؛ فهو يستخدم البيانات كافة. فإذا تجمعت معظم نقاط البيانات على نحو وثيق جدًّا معًا، مع وجود عدد قليل من النقاط النائية، فسيعني ذلك أن الانحراف المعياري صغير. وفي المقابل، إذا كانت نقاط البيانات تتخذ قِيَمًا مختلفة للغاية، حتى إذا كانت تتخذ القِيَم الأكبر والأصغر نفسها، فإن الانحراف المعياري سيكون أكبر بكثير.

(٢-٣) الالتواء

تخبرنا مقاييس التشتت بمدى انحراف القيم المفردة بعضها عن بعض، ولكنها لا تخبرنا بطريقة انحرافها. وبالتحديد لا تخبرنا ما إذا كانت الانحرافات الأكبر تميل إلى أن تكون لدى القيم الكبرى أم القيم الصغرى في مجموعة البيانات. تَذكَّرْ مثالنا عن موظفي الشركة الخمسة، والذي يحصل فيه أربعة موظفين على حوالي ١٠٠٠٠ دولار سنويًّا، بينما يحصل موظف واحد على حوالي عشرة أضعاف ذلك. من شأن أي مقياس للتشتت (الانحراف المعياري على سبيل المثال) أن يخبرنا أن القيم مشتتة على نطاق واسع جدًّا، ولكنه لن يُخبِرنا أن إحدى القيم أكبر بكثير من القيم الأخرى. وبالفعل، فإن الانحراف المعياري

للقيم الخمسة ٩٠٠٠٠ دولار، ٨٩٩٩٩ دولارًا، ٨٩٩٩٨ دولارًا، ٨٩٩٩٧ دولارًا، ١ دولار؛ هو بالضبط نفسه للقيم الخمسة الأصلية. المختلِف هنا هو أن القيمة الشاذة (قيمة ١ دولار) الآن صغيرة جدًّا بدلًا من كونها كبيرة جدًّا. ولرصد هذا الاختلاف، نحتاج إلى إحصائية أخرى لتلخيص البيانات، إحصائية تضع في الاعتبار وتقيس «عدم التناظر» في توزيع القِيَم. يسمَّى أحد أنواع عدم التناظر في توزيع القيم «الالتواء». ويعد مثالنا الأصلي لرواتب الموظفين، الذي يمتلك قيمة واحدة كبيرة على نحو شاذ تبلغ ٩٩٩٩٩ دولارًا، «أيمن الالتواء» (أو موجب الالتواء)؛ لأن توزيع القيم يمتلك «ذيلًا» طويلًا يمتد إلى قيمة واحدة كبيرة للغاية هي ٩٩٩٩٩ دولارًا. لهذا التوزيع العديد من القيم الصغرى وعدد قليل للغاية من القيم الكبرى. وفي المقابل، فإن توزيع القيم المذكور سابقًا، الذي يتضمن شذوذًا عند قيمة ١ دولار، يكون «أيسر الالتواء» (أو سالب الالتواء)؛ لأن الجزء الأكبر من القيم يتراكم معًا، ويوجد ذيل طويل يمتد للأسفل نحو القيمة المفردة الصغيرة جدًّا.

التوزيعات الموجبة الالتواء شائعة كثيرًا، والمثال الكلاسيكي عليها هو توزيع الثروة، والذي يمتلك فيه العديد من الأفراد مبالغ صغيرة فيما يمتلك عدد قليل فحسب من الأفراد مليارات عدة من الدولارات. ويُعَدُّ توزيع رواتب لاعبي البيسبول في الشكل ٢-١ توزيعًا موجب الالتواء بشدة.

(٢-٤) المقاييس التجزيئية

تقدِّم القيم المتوسطة ومقاييس التشتت ومقاييس الالتواء ملخصات إحصائية إجمالية، فتكثف القيم الموجودة في التوزيع إلى أعداد قليلة يسهل التعامل معها. مع ذلك، ربما يكون اهتمامنا مقصورًا على أجزاء فقط من التوزيع؛ على سبيل المثال، ربما نكون مهتمين فحسب بأكبر أو أصغر بضع قيم في مجموعة البيانات؛ مثلًا، أكبر ٥٪ من القيم. الْتَقَيْنا بالفعل الوسيط؛ وهو القيمة التي تكون موجودة في منتصف البيانات؛ بمعنى أن ٥٠٪ من القيم أكبر منها و٥٠٪ أصغر منها. ويمكن تعميم هذه الفكرة؛ فعلى سبيل المثال، «الرُّبَيْع الأعلى» من مجموعة من الأرقام هو القيمة التي يكون ٢٥٪ (أي الربع) من قيم البيانات أكبر منها، أما «الرُّبَيْع الأدنى» فهو القيمة التي يكون ٢٥٪ من قيم البيانات أصغر منها.

وبالمُضِيِّ في هذه التجزئة بدرجة أكبر نجد أن لدينا «العُشَيْرَ» (الذي يقسم مجموعة البيانات إلى أعشار، من العُشَيْر الأدنى وصولًا إلى العُشَيْر الأعلى) و«المُؤَيَّ» (الذي يقسم

البيانات إلى شرائح مئوية). وهكذا يمكن وصف شخص بأنه حقق نتيجة فوق المُؤَيِّ الخامس والتسعين؛ وهذا يعني أنه في أعلى ٥٪ من مجموعة النتائج. والمصطلح العام — الذي يتضمن الرُّبَيْعَ والعُشَيْرَ والمُؤَيَّ وغيرها كحالات خاصة — هو «المقاييس التجزيئية».

الفصل الثالث

جمع بيانات صالحة

البيانات الخام مثل البطاطس الخام؛ عادةً ما تتطلب تنظيفًا قبل الاستخدام.

رونالد إيه ثيستد

توفُّر البيانات نافذة على العالم، ولكن من المهم أن تمنحنا رؤية واضحة. إن النافذة التي تُعانِي من الخدوش أو التشوهات أو وجود علامات على زجاجها من المرجَّح أن تضلِّلنا حيال ما يكمن وراءَها، وينطبق الأمر نفسه على البيانات. فإذا كانت البيانات مشوَّهة أو تالِفة بطريقةٍ ما، يمكن بسهولة أن تنشأ عنها استنتاجات خاطئة. وعمومًا، ليست كل البيانات ذات جودة عالية. في الواقع، يمكنني أن أتعمق أكثر وأشير إلى أنه من النادر أن تقابل مجموعةً من البيانات ليس بها مشاكل في الجودة من أي نوعٍ، ربما إلى حد أنك إذا قابلتَ مجموعة من مثل هذه البيانات «المثالية» فلا بد أن تشك فيها. ربما يجب عليك وقتها أن تسأل عن عمليات الإعداد التي خضعت لها مجموعة البيانات، والتي تجعلها تبدو مثالية. وسوف نعود إلى مسألة الإعداد لاحقًا.

تميل التوصيفات القياسية للأفكار والأساليب الإحصائية الموجودة في الكتب إلى افتراض أن البيانات ليس بها مشاكل (وهنا يصف خبراء الإحصاء البيانات بأنها «نظيفة»، في مقابل البيانات «الملوثة» أو «الفوضوية»). وهذا أمر مفهوم؛ لأن الهدف من هذه الكتب هو وصف الطرق، وينتقص من وضوح الوصف قول ما يجب القيام به إذا كانت البيانات ليست كما ينبغي أن تكون. ومع ذلك، فإن هذا الكتاب مختلف إلى حدٍّ ما؛ فالهدف هنا ليس تعليم آليات الأساليب الإحصائية، وإنما تقديم ونقْل نكهة المجال الحقيقي. ومجال الإحصاء الحقيقي ينبغي أن يتعامل مع البيانات الملوثة.

من أجل توسيع مناقشتنا، نحتاج إلى فهمِ ما يمكن أن تَعنِيه «البيانات الفاسدة»، وكيفية التعرف عليها، وماذا نفعل حيالها. لسوء الحظ، البيانات مثل الناس؛ فيمكن أن «تفسد» بعدد غير محدود من الطرق المختلفة. ومع ذلك، يمكن تصنيف العديد من هذه الطرق على أنها «ناقصة» أو «غير صحيحة».

(١) البيانات الناقصة

تُعَدُّ مجموعة البيانات غيرَ مكتملة إذا كانت بعض الملاحظات غير موجودة، وقد تكون البيانات مفقودة على نحو عشوائي لأسباب لا علاقة لها تمامًا بالدراسة؛ على سبيل المثال، ربما أوقع كيميائي أنبوب اختبار، أو غاب مريض في التجارب السريرية لكريم البشرة عن موعد المتابعة بسبب تأخر الطائرة، أو انتقل شخص من منزله ومِن ثَمَّ لم يمكن الاتصال به من أجل استكمال استبيان المتابعة. ولكن حقيقة أن عنصر بيانات مفقود يمكن أيضًا أن تقدم معلومات في حد ذاتها؛ فعلى سبيل المثال، ربما يرغب الشخص الذي يستكمل استمارة الطلب أو الاستبيان في إخفاء شيءٍ ما، وبدلًا من الكذب الصريح، ربما ببساطة لا يُجِيب عن هذا السؤال. أو ربما أن الأشخاص المعتنقين لوجهة نظر معينة هم فحسب مَن يُجيبون على الاستبيان؛ على سبيل المثال، إذا طُلب من العملاء ملء استمارات تقييم للخدمة التي يتلقَّوْنها، فإن الأشخاص الذين يريدون مناقشة أشياء بخصوص الخدمة ربما يكونون أكثر ميلًا لملء الاستبيان. وإذا لم يُدرَك ذلك في التحليل، فسوف تَنتج صورة مشوهة عن آراء العملاء. واستطلاعات الإنترنت معرَّضة خصوصًا لهذا النوع من العيوب؛ حيث يُكتفى غالبًا بدعوة الناس للإجابة على الاستبيان؛ فلا توجد سيطرة على مدى تمثيل المستجيبين للمجموعة الخاضعة للدراسة بأكملها، أو حتى على احتمالية أن يجيب الأشخاص أنفسهم عدة مرات.

توجد أمثلة أخرى كثيرة لهذا النوع من «التحيز في الاختيار»، ويمكن أن تكون خفية إلى حدٍّ ما؛ على سبيل المثال، من المألوف للمرضى الانسحاب من التجارب السريرية للأدوية. لنفترض أن المرضى الذين شُفُوا أثناء استخدام الدواء لم يعودوا للمقابلة التالية لأنهم شعروا أنها غير ضرورية (بما أنهم قد تعافَوْا). حينها يمكن المسارعة بالاستنتاج بأن الدواء لم ينجح، لأن المرضى الحاضرين هم فقط أولئك الذين لا يزالون مصابين بالمرض.

ظهرت حالة كلاسيكية لهذا النوع من التحيز عندما تنبَّأت جريدة «ليترار ي دايجست» على نحو غير صحيح أن لاندون سيَهزِم روزفلت في الانتخابات الرئاسية في

عام ١٩٣٦ في الولايات المتحدة بأغلبية ساحقة. لسوء الحظ، كانت الاستبيانات قد أُرسلَت فقط للأشخاص الذين لديهم هاتف وسيارة، وفي عام ١٩٣٦ كان هؤلاء الأشخاص أكثر ثراء في المتوسط من إجمالي المجموعة الخاضعة للدراسة. فكان الأشخاص الذين أُرسلت إليهم الاستبيانات لا يمثلون على نحو صحيح كل المجموعة المطلوبة. وكما اتضح، الجزء الأكبر من غيرهم أيَّدوا روزفلت.

ثمة نوع آخر مختلف من حالة الاستنتاجات غير الصحيحة الناشئة عن عدم مراعاة البيانات المفقودة، والذي أصبح حالة إحصائية كلاسيكية ثانوية. هذه الحالة هي حالة مكوك الفضاء «تشالنجر»، الذي انفجر عند إطلاقه في عام ١٩٨٦؛ مما أسفر عن مقتل جميع مَن كانوا على متْنه. في الليلة التي سبقت الإطلاق، عُقد اجتماع لمناقشةِ ما إذا كان ينبغي المضي قدمًا في الإطلاق أم لا؛ حيث إن توقعات درجة الحرارة في موعد الإطلاق أشارتْ إلى أنها منخفضة على نحو كبير. أُنتجت بيانات تبيِّن أنه على ما يبدو لا توجد علاقة بين درجة حرارة الهواء والأضرار التي لحقت ببعض أربطة الصواريخ المساعدة. ومع ذلك، كانت البيانات غير مكتملة، ولم تشمل جميع عمليات الإطلاق التي لم تقع بها «أي» أضرار. كان هذا غير ملائم لأن عمليات الإطلاق التي لم تَقَع فيها أي أضرار أُجريت في الغالب في درجات حرارة أعلى. كان الجدول المحتوي على البيانات «كافة» سيُظهِر علاقة واضحة؛ زيادة احتمالية وقوع الضرر في درجات الحرارة الأقل.

وكمثال أخير، الأشخاص الذين يتقدمون بطلبات للحصول على قروض مصرفية وبطاقات الائتمان، وما شابَهَ ذلك، يجري حساب «مجموع النقاط الائتمانية» لهم؛ وهي تلعب دورًا أساسيًّا في تقدير احتمالية عجْزِهم عن السداد. وتُستمد هذه التقديرات من النماذج الإحصائية المبنية (كما هو موضح في الفصل السادس) باستخدام بيانات من العملاء السابقين الذين سددوا ديونهم بالفعل أو عجزوا عن السداد. ولكن توجد مشكلة؛ فالعملاء السابقون ليسوا ممثلين لجميع الأشخاص الذين تقدموا بطلبات للحصول على قرض. فرغم كل شيء، اختير العملاء السابقون لأنه كان يُعتقد أنهم مخاطرة مأمونة. فلو كان هؤلاء المتقدمون عُدُّوا مخاطرة غير مأمونة في حد ذاتهم وكان من المرجح أن يعجزوا عن السداد، ما كانوا ليُقبَلوا في المقام الأول؛ ومن ثَمَّ لم يكونوا ليدخلوا في البيانات. إن أي نموذج إحصائي لا يأخذ بعين الاعتبار هذا التشويه في مجموعة البيانات من المرجَّح أن يؤدِّي إلى استنتاجات خاطئة. وفي هذه الحالة، يمكن أن يعني هذا انهيار البنك.

إذا كانتْ بعض القِيَم فحسب ناقصة لكل سجل (على سبيل المثال بعض الإجابات على الاستبيان)، يوجد نهجان أساسيان شائعان للتحليل. يتمثل أحد النهجين ببساطة في نبذ أي سجلات غير مكتملة؛ وهذا يتضمن نقطتَي ضعف محتملتين خطيرتين؛ أولاهما: أنه يمكن أن يؤدي لتشوهات يسببها التحيز في الاختيار من النوع الذي نوقش آنفًا؛ فإذا كانت سجلات من نوع معين أكثر عرضة لأن يكون بعض قِيَمها ناقصة، فإن حذف هذه السجلات سوف يترك مجموعة بيانات مشوهة. ونقطة الضعف الخطيرة الثانية هي أنه يمكن أن يؤدي إلى انخفاض هائل في حجم مجموعة البيانات المتاحة للتحليل؛ على سبيل المثال، لنفترض أن استبيانًا يحتوي على مائة سؤال، من الممكن تمامًا ألَّا يُجِيب أي مشارك في الدراسة على «كل» سؤال؛ ومن ثم فإن «جميع» السجلات ستتضمن شيئًا ناقصًا؛ وهذا يعني أن نبذ الردود غير المكتملة من شأنه أن يؤدي إلى نبذ كافة البيانات.

النهج الشائع الثاني لمعالجة القيم الناقصة هو إدخال قيم بديلة؛ على سبيل المثال، لنفترض أن بند العمر ناقص من بعض السجلات، يمكننا حينها استبدال متوسط الأعمار المسجلة بالقيم المفقودة. وعلى الرغم من أن هذا ينتج مجموعة بيانات كاملة (سواء أكملها المشاركون في الدراسة أو أكملناها نحن)، فإنه له عيوب أيضًا؛ ففي هذه الحالة نكون قد اختلقنا البيانات في الأساس.

إذا كان هناك سبب للشك في أن غياب عدد معين إنما يرتبط بالقيمة التي كان سيمتلكها لو كان حاضرًا (على سبيل المثال، إذا كان كبار السن أقل في احتمالية التعريف بسِنِّهم)، فثمة حاجة إلى وجود أساليب إحصائية أكثر تفصيلًا. نحن بحاجة إلى بناء نموذج إحصائي لاحتمالية نقصان البيانات — ربما من النوع الذي يتناوله الفصل السادس — وكذلك للعلاقات الأخرى الموجودة داخل البيانات.

ومن الجدير بالذكر أنه من الضروري قبول حقيقة أنه ليست كل القيم قد سُجِّلَت. ومن الممارسات الشائعة استخدام رمز خاص للإشارة إلى أن القيمة ناقصة؛ على سبيل المثال، من الشائع استخدام رمز N/A اختصارًا لعبارة Not Available بمعنى «غير مُتَاح»، ولكن في بعض الأحيان يتم استخدام رموز رقمية مثل ٩٩٩٩ بالنسبة للعمر. وفي هذه الحالة، الإخفاق في جعل جهاز الكمبيوتر يدرك أن ٩٩٩٩ يمثل القيم الناقصة يمكن أن يؤدي إلى نتيجة غير دقيقة إلى حدٍّ كبير. تَخيَّلْ ما سيكون عليه متوسط العمر المقدَّر عندما يدخل عدد كبير من القيم ٩٩٩٩ في عملية الحساب.

عمومًا، ولعل هذا ينبغي أن يكون متوقَّعًا، لا يوجد حلٌّ مثالي للبيانات الناقصة؛ فجميع طرق التعامل معها تتطلب إقحام نوع من الافتراضات الإضافية، والحل الأفضل هو تقليل المشكلة أثناء مرحلة جمع البيانات.

(٢) البيانات غير الصحيحة

البيانات غير المكتملة هي نوع واحد من مشكلات البيانات، ولكن ربما تكون البيانات «غير صحيحة» بأي عدد من الطرق ولأي عدد من الأسباب. ويوجد مستويات عالية ومنخفضة لأسباب هذه المشكلات.

أحد الأسباب العالية المستوى يتمثل في صعوبة اتخاذ قرار بشأن التعريفات المناسبة (والمتفق عليها عالميًّا). يُعَدُّ معدَّل الجريمة — المشار إليه في الفصل الأول — مثالًا على ذلك، ويُعَدُّ معدل الانتحار مثالًا آخَر. عادةً ما يكون الانتحار نشاطًا فرديًّا؛ لذلك لا يستطيع أحد آخَر أن يعرف على وجه اليقين أنه كان انتحارًا. في أحيان كثيرة تُترك رسالة انتحار، ولكن ليس في جميع الحالات؛ ومن ثم يجب استخلاص دليل على أن الوفاة كانت في الحقيقة انتحارًا. وهذا ينقلنا إلى نطاق غامض؛ لأنه يُثِير مسألة الأدلة ذات الصلة، وعدد الأدلة المطلوبة. علاوة على ذلك، يعمد العديد من المنتحرين إلى إخفاء حقيقة انتحارهم؛ لكي تستطيع الأسرة الحصول على أموال التأمين على الحياة مثلًا.

في موضع مختلف — أكثر تعقيدًا — تتولَّى الوكالة الوطنية لسلامة المرضى في المملكة المتحدة مسئولية وضْع التقارير حول الحوادث التي تقع في المستشفيات، ثم تحاول الوكالة بعد ذلك تصنيفها لتحديد القواسم المشتركة، لكي يمكن اتخاذ الخطوات اللازمة لَمَنْع وقوع الحوادث في المستقبَل. وتكمن الصعوبة في أن الحوادث تُوصَف عن طريق عدة آلاف من الأشخاص المختلفين، وتوصف بطرق مختلفة. وحتى الحادث نفسه يمكن وصفه بأكثر من نحو مختلف جدًّا.

على مستوًى أدنى، غالبًا ما تقع أخطاء في قراءة المقاييس أو تسجيل القيم؛ على سبيل المثال، يوجد اتجاه شائع في قراءة المقاييس وهو التقريب بلا وعي إلى أقرب عدد صحيح؛ فتوزيعات قياسات ضغط الدم المسجَّلة باستخدام مقاييس ضغط الدم القديمة (غير الإلكترونية) تُظهِر اتجاهًا واضحًا لمزيد من القيم المسجَّلة عند ٦٠، ٧٠، ٨٠ ملِّيمترًا من الزئبق من القيم المجاورة، مثل ٦٩ أو ٧٢. وعند أقصى قدْر يمكن أن

تصله أخطاء التسجيل، يمكن أن تُعكس الأرقام (٢٨، بدلًا من ٨٢)، أو يمكن الخلط بين الرقم ٧ المكتوب بخط اليد مع الرقم ١ (وهذا أقل احتمالًا في أوروبا؛ حيث إن ٧ يكتب ٧)، أو قد توضع البيانات في العمود الخطأ في النموذج، وبهذا تتضاعف القيم مصادفة بمقدار عشرة أضعاف، أو ربما يحدث خلط بين النمط الأمريكي لكتابة التاريخ (شهر/يوم/سنة) ونمط المملكة المتحدة (يوم/شهر/سنة)، أو العكس، وما شابه ذلك. في عام ١٧٩٦، طَرَدَ الفلكي الملكي نيفيل ماسكيلين مساعده ديفيد كينبروك على أساس أن مشاهدات الأخير للأوقات التي يَعبُر فيها نجم مختار لخط الزوال عن طريق أحد التلسكوبات في جرينتش لم تكن دقيقة جدًّا. كان هذا الأمر مهمًّا لأن دقة الساعة في جرينتش تتوقف على القياسات الدقيقة لأوقات العبور، وتقديرات خطوط الطول لدى سفن الدولة تعتمد على الساعة، والإمبراطورية البريطانية تعتمد على سفنها. ومع ذلك، فسَّر الباحثون بعد ذلك أسباب عدم الدقة هذه في ضوء تأخُّر رد الفعل النفسي وظاهرة التقريب اللاواعي المذكورة أعلاه. وكمثال أخير من بين كثير من الأمثلة التي كان يمكن أن أختارها، أشار تعداد الولايات المتحدة لعام ١٩٧٠ إلى وجود ٢٨٩ فتاة رُمِّلت وطُلِّقت في الوقت نفسه في سن ١٤. ويجب أن نلاحظ أيضًا نقطة عامة، وهي أنه كلما زاد حجم مجموعة البيانات، زاد عدد المشاركين في تجميعها، وكلما زادت المراحل المشارِكة في معالجتها، زاد احتمال احتوائها على أخطاء.

كثيرًا ما تنشأ أمثلة أخرى لأخطاء البيانات من المستوى الأدنى من وحدات القياس، مثل تسجيل الطول بالمتر بدلًا من القَدَم، أو الوزن بالرطل بدلًا من الكيلوجرام. في عام ١٩٩٩، فُقد «مسبار مناخ المريخ» عندما فشل في دخول الغلاف الجوي للمريخ بالزاوية الصحيحة بسبب الخلط بين قياسات الضغط بوحدَتَي الرطل والنيوتن. وفي مثال آخَر للخلط بين وحدات القياس — وهذه المرة في سياق طبي — كانت مستويات الكالسيوم في الدم عند سيدة مسِنَّة عادةً مستوياتٍ عاديةً، في نطاق ٨٫٦ حتى ٩٫١، لكن بَدَتْ فجأة أنها انخفضت إلى قيمة أقلَّ من ذلك بكثير تبلغ ٤٫٨. كانت الممرضة المسئولة على وشك أن تبدأ في حقنها بالكالسيوم عندما اكتشف الدكتور سلفاتوري بينفينجا أن الانخفاض الظاهري كان ببساطة بسبب أن المختبر غير وحدات القياس التي كان يستخدمها في تقديم تقارير النتائج (من ملِّيجرام لكل ديسيلتر (عُشْر اللتر) إلى ملِّي مكافئ لكل لتر).

(٣) انتشار الخطأ

بمجرد ارتكاب الأخطاء، فإنها يمكن أن تتفشَّى وتسبِّب عواقب وخيمة؛ على سبيل المثال، نُسِبَ عجز الميزانية وتسريح العمال المحتمل في شمال غرب ولاية إنديانا في عام ٢٠٠٦ إلى تأثير خطأ في رقم واحد فقط شقَّ طريقه عبر النظام؛ فأحد المنازل كانت قيمته ١٢١٩٠٠ دولار لكن تغيرت قيمته عن طريق الخطأ إلى ٤٠٠ مليون دولار. وللأسف، استُخدمت هذه القيمة الخاطئة في حساب المعدلات الضريبية.

وفي حالة أخرى، ذكر عدد صحيفة «تايمز» بتاريخ ٢ ديسمبر ٢٠٠٤ كيف أن ٦٦٥٠٠ شركة من حوالي ١٧٠٠٠٠ شركة أُزيلت عن طريق الخطأ من قائمة مستخدمة لتجميع التقديرات الرسمية لناتج البناء في المملكة المتحدة؛ وأدَّى ذلك إلى انخفاض نمو البناء في الربع الأول بنسبة ٢٫٦٪، بدلًا من القيمة الصحيحة التي تقضي بارتفاعه بنسبة ٠٫٥٪؛ وتَرتَّبَ على ذلك أنه في الربع الثاني وَرَدَ أن نسبة النمو تبلغ ٥٫٣٪ بدلًا من النسبة الفعلية البالغة ٢٫١٪.

(٤) الإعداد

كما يجب أن يكون واضحًا من الأمثلة السابقة، فإن عنصرًا أساسيًّا أوليًّا في أي تحليل إحصائي يتمثل في الفحص الدقيق للبيانات والتحقق من وجود الأخطاء وتصحيحها إنْ أمكن. وفي بعض السياقات، يمكن أن تستغرق هذه المرحلة الأولية وقتًا أطول من مراحل التحليل اللاحقة.

ثمة مفهوم رئيسي في تنظيف البيانات هو «القيمة الشاذة». والقيمة الشاذة هي قيمة تختلف كثيرًا عن القيم الأخرى، أو عما هو متوقع، وتكون خارجة عن ذيل التوزيع. وأحيانًا تحدُث هذه القيم المتطرفة بفعل المصادفة؛ فعلى سبيل المثال، على الرغم من أن معظم حالات الطقس تكون معتدلة إلى حدٍّ ما، فإن العواصف الشديدة تحدث بالفعل في بعض الأحيان. ولكن في حالات أخرى ينشأ الشذوذ بسبب أنواع الأخطاء الموضحة سابقًا، مثل مقياس شدة الريح الذي يشير ظاهريًّا إلى عاصفة ضخمة مفاجئة من الرياح في كل منتصف ليل، تزامنًا مع الوقت نفسه الذي يعيد فيه تلقائيًّا معايرة نفسه؛ لذلك يعد البحث عن القيم الشاذة استراتيجية عامة جيدة للكشف عن الأخطاء في البيانات، والتي يمكن بعد ذلك التحقق منها عن طريق شخصٍ ما. وربما تكون هذه القيم قيمًا

شاذة خاصة بمتغيرات مفردة (مثل الرجل البالغ من العمر ٢١٠ سنوات)، أو متغيرات متعددة، ليس أيٌّ منها قيمة شاذة في حد ذاته (مثل الفتاة البالغة من العمر ٥ سنوات ولديها ٣ أطفال).

وبطبيعة الحال، كشْف القيمة الشاذة ليس حلًّا شاملًا للكشف عن الأخطاء في البيانات؛ فرغم كل شيء، يمكن الوقوع في أخطاء تؤدي إلى قيم تظهر طبيعية تمامًا. فربما يُدرَج جنس شخصٍ ما عن طريق الخطأ على أنه أنثى بدلًا من كونه ذكرًا. وأفضل حل هو تبنِّي ممارسات إدخال بيانات تقلل من عدد من الأخطاء. وسنتناول هذا الأمر بالتفصيل في جزء تالٍ.

إذا اكتُشف خطأً واضح، تواجهنا مشكلةُ ما يجب القيام به حياله. يمكن أن نحذف القيمة، معتبرين أنها قيمة ناقصة، ثم نحاول استخدام أحد إجراءات القيم الناقصة المذكورة سابقًا. وأحيانًا يمكننا وضع تخمين ذكي لما كان ينبغي أن تكون عليه هذه القيمة؛ على سبيل المثال، لنفترض أنه خلال تسجيل أعمار مجموعة من الطلاب، حصل الشخص على سلسلة القيم ١٨، ١٩، ١٧، ٢١، ٢٣، ١٩، ٢١٠، ١٨، ١٨، ٢٣. وبدراسة هذه القيم، ربما نعتقد أنه من المرجح أن القيمة ٢١٠ قد دخلت في العمود الخطأ، وأنه ينبغي أن تكون ٢١. وبالمناسبة، لاحظ عبارة «تخمين ذكي» المستخدمة أعلاه. فكما هي الحال مع كل تحليلات البيانات الإحصائية، فإن التفكير المتأني أمر بالغ الأهمية. فليس الأمر مجرد مسألة اختيار طريقة إحصائية معينة وترك الكمبيوتر ليقوم بالعمل؛ فالكمبيوتر لا يقوم إلا بالعمليات الحسابية وحسب.

كان مثال أعمار الطلاب في الفقرة السابقة صغيرًا للغاية؛ إذ كان يحتوي فحسب على عشرة أرقام؛ لذلك كان من السهل النظر فيها وتحديد القيمة الشاذة، ووضع تخمين ذكي حول ما ينبغي أن تكون عليه هذه القيمة. ولكننا نواجه على نحو متزايد مجموعات بيانات أكبر وأكبر. إن مجموعات البيانات المكونة من عدة مليارات من القيم شائعة في الوقت الحاضر في التطبيقات العلمية (مثل تجارب الجسيمات)، والتطبيقات التجارية (مثل الاتصالات)، وغيرها من المجالات الأخرى. وغالبًا ما سيكون مستحيلًا تمامًا استكشاف كل القيم يدويًّا، ويكون علينا أن نعتمد على الكمبيوتر. طوَّر الإحصائيون إجراءات آلية للكشف عن القيم الشاذة، ولكنها لا تحل المشكلة تمامًا. ربما تلفت الإجراءات الآلية الانتباه نحو أنواع معينة من القيم الغريبة، ولكنها ستتجاهل سمات الغرابة التي لم تُخبر عنها. ثم هناك مسألةُ ما يجب القيام به حيال الشذوذ الظاهري

الذي كشفه الكمبيوتر. لا بأس في هذا إذا كان رقمًا واحدًا من هذه المليارِ رقمٍ هو الذي كان موضع شك، ولكن ماذا لو كان مائة ألف رقم في موضع شك؟ مرة أخرى، الفحص والتصحيح عن طريق الأشخاص غير عملي. وللتعامل مع مثل هذه الحالات، طوَّر الإحصائيون مرة أخرى إجراءات آلية، وطورت بعضٌ من أقدم أساليب التحرير والتصحيح الآليةَ تلك في سياق التعدادات والدراسات المسحية الكبيرة، ولكنها دراسات ليست معصومة من الخطأ. خلاصة القول أن الإحصائيين لا يستطيعون — للأسف — صنع المعجزات. إن وجود بيانات رديئة الجودة يجعلنا في خطر الحصول على نتائج رديئة الجودة (بمعنى غير دقيقة وخاطئة وعرضة للخطأ). وأفضل استراتيجية لتجنب ذلك هي الحرص على الحصول على بيانات ذات جودة عالية من البداية.

طورت العديد من الاستراتيجيات لتجنب الأخطاء في البيانات في المقام الأول، وهي تتنوع وفقًا لمجال التطبيق وطريقة جمع البيانات؛ على سبيل المثال، عندما تُنسخ بيانات التجارب السريرية من استمارات سجل الحالة المكتوبة باليد، يوجد احتمال حدوث أخطاء في مرحلة النسخ. وتقلل هذه الأخطاء عن طريق ترتيب تكرار إدخال البيانات مرتين عن طريق شخصين مختلفين يعملان على نحو مستقل، ثم التحقق من وجود أي اختلافات. عند التقدم للحصول على قرض، فإن بيانات الطلب (مثل العمر والدخل والديون الأخرى، وما إلى ذلك) يمكن إدخالها مباشرة إلى جهاز الكمبيوتر، ويمكن لبرامج الكمبيوتر التفاعلية التحقق من الأجوبة بينما يتم إدخالها (على سبيل المثال، إذا كان الشخص مالكًا لمنزل، فهل تشمل ديونه الرهن العقاري؟) وعمومًا، يجب تصميم الاستمارات على نحو يقلل الأخطاء؛ فلا ينبغي أن تكون معقدة على نحو مفرط، ويجب أن تكون جميع الأسئلة واضحة. ومن الواضح أنه من الأفكار الجيدة إجراء دراسة مسحية تجريبية صغيرة للتعرف على أية مشكلات في عملية جمع البيانات قبل الانتقال للتنفيذ الفعلي.

وبالمناسبة، تعد عبارة «خطأ حاسوبي» عبارة مألوفة، ويعد الكمبيوتر كبش فداء شائع عندما تحدث أخطاء في البيانات. ولكن الكمبيوتر يفعل فحسب ما يقال له، مستخدمًا البيانات المُقدَّمَة له. وعندما تحدث الأخطاء، فليس هذا صنيعة يد الكمبيوتر.

(٥) البيانات الرصدية في مقابل البيانات التجريبية

غالبًا ما يكون من المفيد التمييز بين الدراسات «الرصدية» والدراسات «التجريبية»، وبالمثل بين البيانات الرصدية والبيانات التجريبية. تشير الصفة «رصدية» إلى الحالات

التي لا يستطيع المرء فيها أن يتدخل في عملية جمع البيانات؛ فعلى سبيل المثال، في استطلاع حول التوجهات الذهنية للأشخاص حيال السياسيين (انظر أدناه)، تُسأل عينة مناسبة من الأشخاص عن شعورهم، أو في دراسة لخصائص المجرات البعيدة، سوف تخضع هذه الخصائص للرصد والتسجيل. في هذين المثالين، اختار الباحثون ببساطةٍ الأشخاصَ أو الأشياءَ التي سيدرسونها ثم سجلوا خصائص هؤلاء الأشخاص أو الأشياء. لا وجود هنا لفكرة القيام بشيءٍ ما للأشخاص أو المجرات قبل قياسها. في المقابل، في الدراسة التجريبية يتلاعب الباحثون فعليًّا بعناصر الدراسة بطريقة ما؛ على سبيل المثال، في تجربة سريرية ربما يعرِّضون المتطوِّعين لدواء معين قبل أَخْذ القياسات. وفي تجربة تصنيعية لإيجاد الظروف التي تُسفِر عن أقوى منتج نهائي، سيجرِّبون ظروفًا مختلفة.

يتمثَّل أحد الفروق الجوهرية بين الدراسات الرصدية والتجريبية في أن الدراسات التجريبية أكثر فعالية بكثير في تحديد السبب والمسبب؛ على سبيل المثال، ربما نخمن أن طريقة معينة لتعليم الأطفال القراءة (الطريقة «أ» مثلًا) أكثر فعالية من طريقة أخرى (الطريقة «ب»). وفي دراسة وصفية، سوف ننظر للأطفال الذين خضعوا للتعليم باستخدام إحدى الطريقتين ونقارن قدرتهم على القراءة. لكننا لن نكون قادرين على التدخل في توزيع الأطفال الذين يخضعون للطريقة «أ» والذين يخضعون للطريقة «ب»؛ فهذا يتحدد من قِبَل شخص آخَر. يسبب ذلك مشكلة محتملة؛ إذ يعني أنه من الممكن أن توجد اختلافات أخرى بين مجموعتَي تعلُّم القراءة، فضلًا عن طريقة التدريس؛ على سبيل المثال، ولتقديم توضيح صارخ، ربما يُلحِق المدرِّس جميع الأطفال الذين يتعلَّمون على نحو أسرع بالطريقة «أ»؛ أو ربما كان الأطفال أنفسهم مسموحًا لهم بالاختيار، ومال أولئك الأكثر تقدمًا بالفعل في القراءة إلى اختيار الطريقة «أ». إذا كنَّا أكثر تمرسًا في مجال الإحصاء، فربما نستخدم أساليب إحصائية في محاولة للسيطرة على أي اختلافات موجودة مسبقًا بين الأطفال، وكذلك العوامل الأخرى التي نعتقد أنها من المرجَّح أن تؤثِّر على مدى سرعة تعلمهم القراءة. ولكن تظل هناك دائمًا احتمالية وجود تأثيرات أخرى لم نفكِّر فيها، والتي تسبب الفرق.

تتغلب الدراسات التجريبية على هذا الاحتمال عن طريق الاختيار المتعمَّد لكل طفل وللطريقة التي يدرس بها؛ فإذا كنَّا نعرِف بالفعل كل العوامل المُمكِنة، بالإضافة إلى طريقة التدريس — التي يمكن أن تؤثِّر على القدرة على القراءة — يمكننا التأكد من

أن التوزيع على طريقتَي التدريس كان «متوازنًا»؛ على سبيل المثال، إذا كنَّا نظنُّ أن القدرة على القراءة تتأثر بالعمر، يمكننا توزيع العدد نفسه من الأطفال الصغار على كل طريقة. وهكذا، فإن أي اختلافات في القدرة على القراءة ناشئة عن العمر لن يكون لها أي تأثير على الفرق بين مجموعتَينا؛ أي إنه إذا كان للعمر تأثير على القدرة على القراءة، فإن التأثير سيكون نفسه في كلتا المجموعتَين. ومع ذلك، تمتلك الدراسات التجريبية وسيلة أكثر قوة في اختيار أي طفل يخضع لأي طريقة، والتي يُطلَق عليها اسم «التوزيع العشوائي»، وسوف أتناول ذلك فيما يلي:

نتيجة هذا أنه في الدراسة التجريبية يمكن أن نكون أكثر ثقة حيال سبب أي تأثير مرصود. وفي تجربة مقارنة تعليم القراءة، يمكننا أن نكون أكثر ثقة أن أي فرْق في القدرة على القراءة بين المجموعتَين هو نتيجة لطريقة التعليم، وليس نتيجة عامل آخَر.

للأسف، ليس من الممكن دائمًا إجراء التجارب بدلًا من الدراسات الرصدية. فلا يمكننا مثلًا تعريض المجرات المختلفة لظروف مختلفة! وعلى أي حال، ربما يكون من المضلِّل في بعض الأوقات استخدام المنهج التجريبي؛ ففي كثير من الدراسات المسحية الاجتماعية، يتمثل الهدف في معرفة حال السكان الحقيقي، لا في «ماذا سيكون التأثير الناتج إذا فعلنا كذا وكذا؟» ومع ذلك، إذا كنَّا نريد بالفعل أن نعرف ماذا سيكون تأثير أي تدخُّل محتمل، فإن الدراسات التجريبية تُعَدُّ استراتيجية أفضل. هذا النوع من الدراسات واسع الانتشار في قطاع الصناعات الدوائية والطب وعلم النفس، ومجال التصنيع والصناعات التحويلية، كما يُستخدم على نحو متزايد في تقييم السياسة الاجتماعية وفي مجالات مثل إدارة قيمة العملاء.

وعمومًا، عند جمع البيانات بهدف إجابة أو استكشاف بعض الأسئلة، كلما زادت البيانات التي تُجمَع، زادت دقة الإجابة التي يمكن الحصول عليها؛ وهذا نتيجة لـ «قانون الأعداد الكبيرة»، الذي سيناقَش في الفصل الرابع. ولكن جمع المزيد من البيانات يفرض تكلفة أكبر. ولذلك فمن الضروري التوصل إلى حلٍّ وسط مناسب بين كمية البيانات التي تُجمَع وتكلفة جمعها. تقبع تخصصات فرعية متعددة من الإحصاء في قلب هذه العملية، وعلى وجه الخصوص، يُعَدُّ «التصميم التجريبي» و«مسح العينات» نوعين من التخصصات الرئيسية.

(٦) التصميم التجريبي

رأينا بالفعل أمثلة لتجارب بسيطة جدًّا. وتتمثل إحدى أبسط التجارب في تجربةٍ سريريةٍ ثنائيةِ المجموعةِ تَستخدم عينات عشوائية. والهدف هنا هو المقارنة بين علاجين من العلاجات البديلة («أ» و«ب»، مثلًا) لكي نستطيع معرفة أيهما ينبغي إعطاؤه لمريض جديد. ولاكتشاف ذلك، نقدم العلاج «أ» إلى إحدى عينتَيِ المرضَى، والعلاج «ب» إلى العينة الأخرى من المرضى، ونُقَيِّم فعالية العلاج. وإذا تفوَّق «أ» على «ب» في المتوسط، فإننا سنوصي أن يتلقَّى المريض الجديد العلاج «أ». سيعتمد معنى كلمة «تفوق» في الجملة السابقة على الدراسة نفسها؛ إذ يمكن أن تعني «يشفي المزيد من المرضى»، أو «يطيل متوسط العمر»، أو «يسبب متوسط انخفاض أكبر في الألم»، أو ما إلى ذلك.

كما لاحظنا بالفعل سابقًا، إذا كانت مجموعتا المرضى تختلفان على نحوٍ ما، فإن الاستنتاجات التي يمكن أن نستخلصها محدودة. فإذا كان المرضى الذين تلقَّوا العلاج «أ» جميعًا من الذكور، والذين تلقَّوا العلاج «ب» كلهم من الإناث، فإننا لن نعرف ما إذا كان أي فرق بين المجموعتَين لاحظناه يرجع إلى العلاج أم إلى اختلاف الجنس؛ إذ ربما تتحسن صحة الإناث أسرع بغضِّ النظر عن العلاج. وتنطبق النقطة نفسها على أي عامل آخَر؛ العمر أو الطول أو الوزن أو مدة المرض أو تاريخ العلاجات السابقة، أو ما إلى ذلك.

إحدى استراتيجيات تقليل هذه الصعوبة تكمن في توزيع المرضى عشوائيًّا على مجموعتَي العلاج. تكمن قوة هذا النهج في أنه على الرغم من عدم ضمانه لوجود توازن (على سبيل المثال، من الممكن أن تؤدي عملية التوزيع العشوائي إلى وجود نسبة أعلى بكثير من الذكور في مجموعة واحدة عن الأخرى)، فإن القواعد الأساسية للاحتمال (التي نناقشها في الفصل الرابع) تخبرنا أن اختلالات التوازن الكبيرة غير مرجحة الحدوث على نحو كبير. في الواقع، من الممكن التعمق أكثر من هذا بحساب احتمالية حدوث درجات عدم التوازن المختلفة. وهذا بدوره يُتيح لنا حساب مدى الثقة التي يجب أن نمتلكها حيال استنتاجاتنا.

وعلاوة على ذلك، إذا كان التوزيع العشوائي «مزدوجَ التعمية»، فلا يوجد خطر التحيز اللاواعي الذي يتدخل في عملية التوزيع أو قياس المرضى. وتكون الدراسة مزدوجةَ التعمية إذا لم يكن المريض ولا الطبيب الذي يُجرِي التجربة يعرف أي علاج يتلقَّاه المريض. ويمكن تحقيق ذلك عن طريق جعل الأقراص أو الأدوية تبدو متطابقة،

وترميزها ببساطة بالحرفين «س» و«ص» دون الإشارة إلى ماهية العلاج. وفي وقت لاحق فحسب — بعد أن يكشف التحليل أن «س» أفضل من «ص» — تُفَك شفرة الترميز، لتوضيح أن «س» هو في حقيقته العلاج «أ» أو «ب».

إن التجربة السريرية الثنائية المجموعة التي تَستخدم عيناتٍ عشوائيةً بسيطةٌ جدًّا، ولها صور موسعة واضحة؛ فعلى سبيل المثال، يمكننا التوسع فيها على الفور إلى أكثر من مجموعتَي علاج. ومع ذلك، من أجل التنويع، سوف أغيِّر الأمثلة. يرغب مُزارع في معرفة أيٍّ من مستويات الماء المنخفضة أو العالية أفضل فيما يخص إنتاج غلة أكبر من المحاصيل. يمكنه إجراء تجربة بسيطة ثنائية المجموعة من النوع المذكور سابقًا لتحديد هذا. وبما أننا نعلم أن النتائج ليست متوقَّعة تمامًا، فسوف يريد تعريض أكثر من صوبة واحدة لمستوى مياهٍ منخفِض، وأكثر من صوبة واحدة لمستوًى عالٍ، ثم يحسب متوسط الغَلَّة في كل مستوًى؛ فعلى سبيل المثال، ربما يقرر استخدام أربع صوبات لكل مستوًى. وهذا هو بالضبط نوع التصميم نفسه كما في دراسة طرق التدريس السابقة.

ولكن لنفترض الآن أن المزارع يريد أيضًا أن يعرف أيٌّ من مستويات الأسمدة المنخفضة والعالية أكثر فعالية. الشيء البديهي القيام به هو إجراء تجربة أخرى ثنائية المجموعة؛ هذه المرة باستخدام أربع صوبات تتلقَّى مستوًى منخفضًا من الأسمدة وأربعة تتلقَّى مستوًى عاليًا. هذا جيد جدًّا، ولكن الإجابة على كلا السؤالين — السؤال عن مستوى الماء والسؤال عن مستوى الأسمدة — تتطلب ما مجموعه ست عشرة صوبة. وإذا كان المزارع مهتمًّا أيضًا بفعالية مستويات الرطوبة المنخفضة والعالية، ودرجة الحرارة، وساعات التعرض لضوء الشمس، وما شابه ذلك؛ فسنرى أنه سيكون بحاجة لعدد كبير للغاية من الصوبات.

توجد طريقة ذكية للغاية للالتفاف حول ذلك؛ وهي باستخدام مفهوم تصميم التجارب «العاملية»؛ فبدلًا من تنفيذ تجربتين منفصلتين، واحدة للمياه وواحدة للأسمدة، يستطيع المزارع معالجة صوبتين باستخدام «أسمدة منخفضة، وماء منخفض»، واثنتين «منخفضة، عالٍ»، واثنتين «عالية ومنخفض»، واثنتين «عالية، عالٍ». هذا يتطلب فقط ثماني صوبات زراعية، ومع ذلك نظل نعالج أربعًا منها بمستوى مياه منخفض وأربعًا بمستوى مياه عالٍ، وكذلك أربع صوبات بمستوى أسمدة منخفض وأربعًا بمستوى أسمدة عالٍ؛ ومن ثم فإن نتائج التحليل سوف تكون دقيقة تمامًا كما لو كنَّا أجرَيْنا تجربتين منفصلتين.

في الواقع، يمتلك هذا التصميم العاملي (المياه والأسمدة كلاهما «عامل») ميزة إضافية جذابة؛ فهو يُتيح لنا معرفةَ ما إذا كان تأثير مستوى السماد مختلفًا عند مستويَيِ المياه؛ فربما يختلف الفرق بين المحصول مع مستويَيِ الأسمدة المنخفِض والعالي في حالة اختلاف مستوى المياه. وهذا يُسمَّى «تأثير التفاعل»، ولا يمكن فحْصُه في نهْج إجراء تجربتين منفصلتين.

جرى التوسع في هذه الفكرة الأساسية بطُرُق عديدة لإنتاج أدوات قوية للغاية للحصول على معلومات دقيقة من أجْل الوصول للحدِّ الأدنى من التكلفة. وعند ضمِّها إلى غيرها من أدوات التصميم التجريبي، مثل التوازن والتوزيع العشوائي والسيطرة على التأثيرات المعروفة، نتجت بعض التصاميم التجريبية المتطورة للغاية.

أحيانًا في التجارب تكون الأمور غير الإحصائية مهمَّة؛ فعلى سبيل المثال، في التجارب السريرية والدراسات الطبية ودراسات السياسة الاجتماعية الأخرى، ربما تكون الأمور الأخلاقية ذات صلة؛ ففي تجربة سريرية تقارِن علاجًا جديدًا مقترَحًا مع علاج وهمي (غير نشِط)، سنكون على معرفة بأنَّ نِصْف المرضى المتطوِّعين سيتلقَّوْن شيئًا ليس له أي تأثير بيولوجي. هل هذا مناسب؟ هل يوجد خطر أن يُعانِي أولئك الذين يتناولون العلاج الجديد المقترح من آثار جانبية؟ مثل هذه الأشياء يجب أن تكون متوازنة مع حقيقة أن أعدادًا لا تُحصَى من المرضى في المستقبَل سوف يستفيدون ممَّا يتم معرفته خلال التجربة.

(٧) مسح العينات

تخيل أنه من أجل إدارة البلاد على نحو فعال، نودُّ أن نعرف متوسط الدخل لمليون شخص عامل من الرجال والنساء في بلدة معينة. ظاهريًّا، يمكننا تحديد هذا عن طريق سؤالِ كلٍّ منهم عن دخله، وحساب متوسط النتائج. أما عمليًّا، فإن هذا سيكون صعبًا للغاية، ويكاد يكون مستحيلًا. وفضلًا عن أي شيء آخر، من المرجَّح أن يتغير الدخل على مدى الوقت الذي سيستغرقه جمع البيانات؛ فربما يترك بعض الناس وظائفهم أو يغيرونها، وربما يتلقَّى البعض الآخر علاوات، وما إلى ذلك. وعلاوة على ذلك، فإن تعقُّب كل شخص سيكون مكلِّفًا للغاية. ربما نحاول خفض التكاليف من خلال الاعتماد على الهاتف، لا المقابلات الشخصية، ولكنْ كما رأَيْنا سابقًا في الحالة المتطرِّفة للانتخابات

الرئاسية في الولايات المتحدة لعام ١٩٣٦، يوجد خطر كبير بأننا سوف نَغفل عن شرائح مهمة من السكان.

ما نحتاجه هو طريقةٌ ما لتقليل تكلفة جمع البيانات لكنها في نفس الوقت تجعل العملية أسرع، وتجعلها — إذا أمكن — أكثر دقة أيضًا. بصياغة الأمر بهذه الطريقة، ربما يبدو الأمر كأنه مهمة شاقَّة، ولكن الأفكار والأدوات الإحصائية التي تتمتع بهذه الخصائص موجودة. والفكرة الرئيسية هي فكرة قابلناها عدة مرات من قبلُ؛ وهي فكرة العيِّنة.

لنفترض أنه بدلًا من معرفةِ ما يحصل عليه كل واحد من المليون موظف، سألنا ببساطةٍ ألفَ موظف منهم. لكن علينا بوضوح الآن أن نكون حَذِرين بشأن الألف موظف الذين نسألهم بالضبط. وأسباب ذلك هي في الأساس الأسباب نفسها التي دعتْنا عندما كنَّا نصمِّم التجربة الثنائية المجموعة البسيطة إلى اتخاذ خطوات لضمان أن الفرق الوحيد بين المجموعتَين كان أن واحدة تتلقَّى العلاج «أ» والأخرى تتلقَّى العلاج «ب»؛ لذا علينا الآن أن نتأكد أن الأشخاص الألف المحددين الذين نتواصل معهم يمثلون الموظفين المليون على نحو تام.

ما الذي نعنيه بكلمة «ممثِّل»؟ على نحو مثالي، ينبغي أن تكون عيِّنتنا المكوَّنة من ألف موظف تحتوي على نسبة الرجال نفسها الموجودة في المجموعة الكاملة الخاضعة للدراسة، والنسبة نفسها من الشباب، والنسبة نفسها من العاملين بدوام جزئي، وما إلى ذلك. نستطيع ضمان ذلك إلى حدٍّ ما من خلال اختيار ألف موظف بحيث تكون نسبة الرجال — على سبيل المثال — صحيحة. ولكن من الواضح أنه يوجد قيد عملي لما يمكننا موازنته عمدًا بهذه الطريقة.

شاهدنا كيفية التعامل مع هذه الصعوبة عندما تناولنا التصميم التجريبي؛ وذلك من خلال «التوزيع العشوائي» للمرضى على كل مجموعة من المجموعتَين. في حالتنا هذه سنتعامل معها عن طريق «أخذ عينة عشوائية» من ألف شخص من مجموعة الموظفين الكلية الخاضعة للدراسة. ومرة أخرى، رغم أن هذا لا يضمن أن العينة ستكون مشابِهة في تكوينها للمجموعة الخاضعة للدراسة، فإن الاحتمالية الأساسية تُخبِرنا أن فرصة الحصول على عيِّنة مختلِفة كثيرًا ضئيلة جدًّا. وتحديدًا، يترتب على ذلك أن احتمالية أن تكون تقديراتنا لمتوسط الدخل، والمستمَدَّة من العينة، مختلفةً كثيرًا عن متوسط الدخل في المجموعة الخاضعة للدراسة بأكملها؛ ضعيفةٌ للغاية. وفي الواقع، ثمة خاصيتان

للاحتمالات — سوف نتناولهما في وقت لاحق؛ هما «قانون الأعداد الكبيرة» و«مبرهنة النهاية المركزية» — تُخبِرانِنَا أيضًا أننا يمكننا جعل هذه الاحتمالية ضئيلة كما نشاء من خلال زيادة حجم العينة. ويتضح لنا أن ما يهم حقًّا ليس مدى كِبَر نسبة العينة إلى المجموعة الكلية، وإنما ببساطة مدى كبر حجم العينة. فسيكون تقديرنا — المستنِد إلى حجم عينة مكونة من ألف شخص — بالدقة نفسها إذا كانت المجموعة الكلية الخاضعة للدراسة تتألف من عشرة ملايين أو عشرة مليارات شخص. وبما أن حجم العينة يرتبط ارتباطًا مباشرًا بتكلفة جمع البيانات، فإن لدينا الآن علاقة مباشرة بين الدقة والتكلفة؛ فكلما كان حجم العينة أكبر، زادت التكلفة، ولكن قلَّ احتمال الانحراف الكبير بين تقدير العينة ومتوسط المجموعة الكلية الخاضعة للدراسة.

في حين أن «أخذ عينة عشوائية مكونة من ألف شخص من المجموعة الكلية» للعاملين في المدينة قد يبدو كأنه عملية بسيطة، فإنها في واقع الأمر عملية تتطلب عناية شديدة؛ فعلى سبيل المثال، لا يمكننا ببساطةٍ اختيارُ ألف شخص من أكبر شركة في المدينة؛ لأن هذه العينة قد لا تكون ممثلة للعاملين المليون جميعهم. وبالمثل، لا يمكننا الاتصال بعينة عشوائية من بيوت الأشخاص في الساعة الثامنة مساءً؛ لأننا سنغفل عن أولئك الذين يعملون في وقت متأخر، وربما يختلف هؤلاء العمال في متوسط الدخل عن الآخرين. وعمومًا، للتأكد من أن العينة المكونة من ألف شخص مُمثِّلة على نحو مناسب للمجموعة الكلية، فإننا بحاجة إلى «إطار المعاينة»؛ وهو قائمة تضم المليون العاملين جميعهم في المجموعة الخاضعة للدراسة، والتي يمكن أن نختار منها ألف شخص عشوائيًّا. إن وجود مثل هذه القائمة يضمن أن احتمالية تضمين كل الأشخاص في العينة متساوية.

تُعَدُّ فكرة «أخذ العينات العشوائية البسيطة» هذه أساسًا لعملية مسح العينات؛ فقد شكَّلْنا إطار معاينة، ومنه نختار عشوائيًّا الأشخاص الذين سنُضمِّنهم في عيِّنتنا، ثم نتعقَّبهم (من خلال مقابلة شخصية أو اتصال هاتفي أو رسالة أو بريد إلكتروني، أو بأي طريقة) ونسجل البيانات التي نريدها. وقد طُورت هذه الفكرة الأساسية بالعديد من الطرق الدقيقة والمتطورة جدًّا؛ مما أسفر عن نهج أكثر دقة وأقل تكلفة؛ على سبيل المثال، إذا كنَّا ننوي مقابلة كل شخص من الألف المشاركين في الدراسة، فإن ذلك يمكن أن يكون مكلِّفًا جدًّا من حيث الوقت ونفقات التنقل. سيكون من الأفضل — من هذا المنظور — اختيار المشاركين في الدراسة من عناقيد جغرافية محلية صغيرة. ويوسع «أخذ العينات العنقودية» عملية أخذ العينات العشوائية البسيطة من خلال السماح

بذلك؛ فبدلًا من اختيار ألف شخص من المجموعة الخاضعة للدراسة عشوائيًّا، فإن هذا النهج يختار (مثلًا) عشر مجموعات تتكون كلٌّ منها من مائة شخص، بحيث يعيش كل الأشخاص في كل مجموعة بعضهم بالقرب من بعض. وبالمثل، يمكننا التأكد من تحقيق التوازن في بعض العوامل، بدلًا من مجرد الاعتماد على إجراء أخذ العينات العشوائية، إذا فرضنا التوازن على طريقة اختيار العينة؛ على سبيل المثال، يمكننا أن نختار عشوائيًّا عددًا من النساء من المجموعة الخاضعة للدراسة، ونختار عشوائيًّا على نحو منفصل عددًا من الرجال من المجموعة الخاضعة للدراسة؛ حيث يتم اختيار الأعداد بحيث تكون نسب الذكور والإناث هي نفسها كما هي الحال في المجموعة الإجمالية الخاضعة للدراسة. يُعرف هذا الإجراء بأنه «الطريقة الشرائحية لأخذ العينات»؛ لأنه يقسِّم المجموعة الكلية الخاضعة للدراسة المدرجة في إطار العينة إلى شرائح (الرجال والنساء في هذه الحالة). وإذا كان المتغير المستخدم في الشرائح (الجنس في هذا المثال) يرتبط ارتباطًا قويًّا بالمتغير الذي نهتم به (الدخل هنا)، يمكن أن يسفر هذا عن تحسن في الدقة لحجم العينة نفسه.

وعمومًا، في عملية المعاينة، نكون محظوظين للغاية إذا حصلنا على ردود من جميع الأشخاص الذين نتواصل معهم. يوجد دائمًا مقدار من عدم الاستجابة، وهذا يعود بنا إلى مشكلة البيانات الناقصة التي ناقشناها سابقًا، وكما رأينا، يمكن للبيانات الناقصة أن تؤدي إلى عينة متحيزة واستنتاجات غير صحيحة. فإذا رفض الذين يحصلون على رواتب كبيرة المشاركة في الدراسة، فسوف نبخس تقدير متوسط الدخل في المجموعة الخاضعة للدراسة. وبسبب هذا، طور خبراء الدراسات المسحية مجموعة متنوعة من وسائل تقليل وضبط عدم الاستجابة، بما في ذلك تكرار التواصل مع غير المستجيبين وإجراءات إعادة التقييم الإحصائي.

خاتمة

تناول هذا الفصل المواد الخام للإحصائيات؛ وهي البيانات. وقد صيغت تقنيات جمع بيانات متطورة على يد الإحصائيين للحصول على أقصى قدر من المعلومات بالحد الأدنى من التكلفة، ولكن سيكون من السذاجة الاعتقاد بأنه يمكن عادة الحصول على بيانات مثالية. إن البيانات انعكاس للعالَم الحقيقي، والعالَم الحقيقي معقَّد. وإدراكًا لهذا، طور الإحصائيون أيضًا أدوات للتعامل مع البيانات ذات الجودة الرديئة. ولكنْ من المهم أن نُدرِك أن الإحصائيين ليسوا سَحَرة. وينطبق القول المأثور القديم: «مُدخَلات عديمة النفع تساوي نتائج عديمة النفع» تمامًا على الإحصائيات كما ينطبق على كل شيء آخر.

الفصل الرابع

الاحتمالات

كونك إحصائيًّا يعني أنك لن تُضطر أبدًا إلى القول بأنك متأكد.

مجهول

(١) جوهر المصادفة

أحد التعريفات التي قدمت في الفصل الأول حول الإحصاء هو أنه علم التعامل مع عدم اليقين. وبما أنه من الواضح للغاية أن العالم مليء بعدم اليقين، فإن هذا أحد أسباب هيمنة الأفكار والأساليب الإحصائية. إن المستقبل مجهول ولا نستطيع أن نكون واثقين بشأن ما سيحدث. وبالفعل يحدث ما هو غير متوقَّع؛ فتتعطل السيارات ونقع في حوادث ويضرب البرق، وخشية أن أقدم انطباعًا بأن كل الأمور سيئة، أقول إن هناك مَن يفوزون حتى باليانصيب. وفي أبسط الحالات، نحن لا نعلم يقينًا أي حصان سيفوز بالسباق أو أي عدد سوف يَظهَر عند إلقاء نرد. وفوق ذلك كله، لا نستطيع التنبؤ بطول الحياة التي سنعيشها.

لكن على الرغم من كل هذا، يتمثل أحد أعظم الاكتشافات التي توصلت إليها البشرية في أنه يوجد مبادئ معينة تحكم سَيْر المصادفة وعدم اليقين. ربما يبدو هذا تناقضًا في المصطلحات؛ فالأحداث غير اليقينية بطبيعتها لا تنطوي على يقين؛ فكيف إذن توجد قوانين طبيعية تحكم سير هذه الأمور؟

إحدى الإجابات على هذا السؤال هي أنه في حين أن الأحداث الفردية ربما تكون غامضة وغير قابلة للتنبؤ بها، فإنه غالبًا ما يكون من الممكن الخروج بتعميم ينطبق على مجموعة من الأحداث. المثال الكلاسيكي لذلك هو إلقاء العملة؛ فرغم أنني لا أستطيع أن

أقول ما إذا كانت العملة ستظهر وجه الصورة أم الكتابة بعد أي عملية إلقاء منفردة، يمكنني أن أقول بثقة كبيرة إنه إذا أُلقيت العملة عدة مرات فإنها ستظهر وجه الصورة في حوالي نصف عدد المرات ووجه الكتابة في حوالي نصف عدد المرات. (أفترض هنا أن العملة «عملة متزنة»، وأنه لا تُستخدم أي خدعة بالأيدي أثناء إلقائها.) وثمة مثال آخر في هذا النطاق هو تحديد ما إذا كان المولود ذكرًا أم أنثى؛ فتحديد الجنس خلال عملية الحمل أمر خاضع للمصادفة البحتة ولا يمكن التنبؤ به. ولكننا نعرف أنه على مدار العديد من حالات الولادة فإن أكثر من نصف عدد المواليد بقليل سيكونون ذكورًا.

تُعَدُّ هذه السمة الطبيعية القابلة للملاحظة مثالًا للقوانين التي تحكم عدم اليقين، ويطلق عليها اسم «قانون الأعداد الكبيرة» بسبب حقيقة أن النسبة تقترب أكثر وأكثر من قيمة معينة (النصف في حالات العملة المتزنة ونوع جنس المواليد) كلما زاد عدد الحالات التي ننظر فيها. لهذا القانون تبعات متعددة، وهو واحد من أقوى الأدوات الإحصائية في ترويض عدم اليقين والسيطرة عليه والسماح لنا بالاستفادة منه. وسنعود إليه لاحقًا في هذا الفصل، وعلى نحو متكرر خلال الكتاب.

(٢) فهم الاحتمالات

لكي نتمكن من مناقشة مسائل عدم اليقين وعدم القدرة على التنبؤ دون غموض، فإن علم الإحصاء يستخدم — مثل أي علم آخَر — لغة دقيقة؛ وهي لغة «الاحتمالات». وإذا كان هذا هو أول تعرُّض لك للغة الاحتمالات، إذن يجب أن أحذرك من أنك سوف تكون بحاجة لبذل بعض الجهد من أجل فهمها، كما هي الحال مع أول تعرض للمرء لأي لغة جديدة. وبوضع ذلك في الاعتبار، ربما تجد في الواقع أن هذا الفصل يتطلب القراءة أكثر من مرة واحدة؛ فربما ترغب في إعادة قراءة هذا الفصل مرة أخرى عندما تصل إلى نهاية الكتاب.

ازدهر تطور لغة الاحتمالات في القرن السابع عشر. وقد أرسى قواعدها علماء الرياضيات أمثال بليز باسكال وبيير دي فيرما وكريستيان هيجنز وجاكوب برنولي، ومن بعدهم بيير سيمون لابلاس وأبراهام دي موافر وسيميون-دنيس بواسون وأنطوان كورنو وجون فين، وغيرهم. وبحلول أوائل القرن العشرين، كانت كل الأفكار اللازمة لعلم احتمالات قوي متوافرة. وفي عام ١٩٣٣، قدَّم عالِم الرياضيات الروسي أندريه

كولموجوروف مجموعة من البديهيات التي قدمت «حسابًا» رياضيًّا رسميًّا كاملًا للاحتمالات. ومنذ ذلك الحين، اعتُمد نظام البديهيات هذا عالميًّا تقريبًا.

توفِّر بديهيات كولموجوروف آليةً يمكن من خلالها التعامل مع الاحتمالات، لكنها بنية رياضية. ولاستخدام هذه البنية لتقديم بيانات حول العالم الحقيقي، من الضروري الإشارة إلى ما تمثِّله الرموز الموجودة في الآلية الرياضية الموجودة في هذا العالم؛ أي إننا بحاجة إلى قول ما «تعنيه» الرياضيات.

يعين حساب الاحتمالات أرقامًا بين ٠ و١ للأحداث غير المؤكدة لتمثيل احتمالية حدوثها. يعني الاحتمال ١ أن هذا الحدث مؤكد (على سبيل المثال، احتمال أنه لو أن أحدهم نظر من نافذة حجرة مكتبي بينما كنتُ أكتب هذا الكتاب، لرآني جالسًا إلى مكتبي). والاحتمال ٠ يعني أن الحدث مستحيل (على سبيل المثال، احتمال أن شخصًا ما سوف ينهي سباق ماراثون في عشر دقائق). وبالنسبة لحدثٍ ما «يمكن» أن يحدث ولكنه ليس مؤكدًا ولا مستحيلًا، فإن رقمًا بين ٠ و١ يمثل «احتمال» حدوثه.

إحدى طرق النظر إلى هذا الرقم هي أنه يمثل «درجة اعتقاد» المرء أن الحدث سوف يحدث. سوف يمتلك الأشخاص المختلفون معلومات أكثر أو أقل متعلقة بكون الحدث سيقع أم لا؛ لذلك ربما يُتوقع أن يمتلك الأشخاص المختلفون درجات مختلفة من الاعتقاد؛ وهذا يعني احتمالات مختلفة لهذا الحدث. ولهذا السبب، تُسمَّى وجهة النظر تلك حيال الاحتمال الاحتمالَ «الذاتي» أو «الشخصي»؛ فهي تعتمد على مَن يقيِّم الاحتمال. ومن الواضح أيضًا أن الاحتمال لدى الشخص ربما يتغير مع توافر المزيد من المعلومات. فربما تبدأ باحتمال — درجة اعتقاد — تبلغ ١ / ٢ أن عملة معينة سوف تستقر ووجه الصورة لأعلى (على أساس تجربتك السابقة مع قذف عملات معدنية أخرى)، ولكن بعد مراقبة استقرار العملة ووجه الصورة لأعلى ١٠٠ مرة متتالية دون استقرارها على وجه الكتابة قط، ربما تصبح متشككًا وتغير احتمالاتك الشخصية بأن تستقر هذه العملة على وجه الصورة لأعلى.

وقد طُورت أدوات لتقدير الاحتمالات الذاتية للأفراد على أساس استراتيجيات المراهنة، ولكن كما هي الحال مع أي إجراء للقياس، ثمة قيود عملية على مدى دقة تقدير الاحتمالات.

تتمثل وجهة نظر مختلفة لاحتمالات وقوع حدثٍ ما في أنها عدد مرات وقوع هذا الحدث إذا تكررت الظروف على نحو متطابق لعدد لا نهائي من المرات. ويُعَدُّ مثال

قذف العملة المتزنة السابق توضيحًا لهذا؛ فقد رأينا أنه بينما تقذف العملة، فإن نسبة ظهور الصورة تقترب أكثر وأكثر من قيمة محددة. وتعرَّف هذه القيمة على أنها احتمال استقرار العملة على وجه الصورة لأعلى في أي عملية قذف واحدة. ونظرًا لدور التكرارات، أو عدد المرات، في تحديد هذا التفسير للاحتمالات، فإنه يسمى التفسير «التكراري».

وتمامًا كما هي الحال مع النهج الذاتي، توجد قيود عملية تمنعنا من إيجاد الاحتمالات التكرارية بالضبط؛ فعمليتا قذفٍ لعملةٍ ما لا يمكن أن تمتلكا حقًّا ظروفًا متطابقة تمامًا؛ فسوف تبلى بعض الجزيئات من العملة في الرمية الأولى، وستختلف تيارات الهواء، وسترتفع درجة حرارة العملة قليلًا جراء التَّماسِّ مع الأصابع في المرة الأولى. وعلى أي حال سيكون علينا وقف قذف العملة في وقتٍ ما؛ لذلك لا يمكننا قذفها فعليًّا لعدد لا نهائي من المرات.

هذان التفسيران المختلفان لما تعنيه الاحتمالات لهما خصائص مختلفة. فيمكن استخدام النهج الذاتي لتعيين احتمال معين لحدثٍ فريد من نوعه؛ حدثٍ يكون التفكير في تكراره في ظل ظروف مماثلة لعدد لا نهائي — أو حتى عدد كبير — من المرات لا معنى له؛ على سبيل المثال، ليس هناك معنًى لاقتراح عمل سلسلة لا نهائية من المحاولات المتطابقة لاغتيال الرئيس المقبل للولايات المتحدة، بحيث يؤدي بعضها لنتيجةٍ ما والبعض الآخر لنتيجة أخرى؛ لذلك يبدو من الصعب تطبيق التفسير التكراري على مثل هذا الحدث. من ناحية أخرى، فإن النهج الذاتي ينقل الاحتمالات من كونها خاصية موضوعية للعالم الخارجي (مثل الكتلة أو الطول) إلى كونها خاصية للتفاعل بين الراصد والعالم؛ فالاحتمالات الذاتية تجعل الراصد هو الأساس. قد يشعر البعض أن هذا نقطة ضعف؛ فهذا يعني أن الأشخاص المختلفين يمكنهم استخلاص استنتاجات مختلفة من التحليل نفسه للبيانات نفسها. وقد يعتبره البعض الآخر نقطة قوة؛ إذ إن الاستنتاجات تتأثر بمعرفتك السابقة.

مع ذلك، توجد تفسيرات أخرى للاحتمال؛ فعلى سبيل المثال، يفترض النهج «الكلاسيكي» أن جميع الأحداث تتكون من مجموعة من الأحداث الابتدائية المتساوية الاحتمال؛ فعلى سبيل المثال، رمي النرد قد ينتج الرقم ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦، وتماثل النرد يشير إلى تَساوي احتمالية ظهور هذه النتائج الست، وهكذا كل رقم لديه احتمال يبلغ ١ / ٦ (يجب أن يكون مجموعها ١، نظرًا لأنه من «المؤكد» أن واحدًا من الأرقام ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ سوف يظهر). واحتمال الحصول على عدد

زوجي — على سبيل المثال — هو مجموع الاحتمالات المتساوية لكل أحداث الحصول على ٢ أو ٤ أو ٦؛ ومن ثَمَّ فهو يساوي ١ / ٢. ومع ذلك، في ظروف أقل اصطناعية، توجد صعوبات في تحديد ماهية هذه الأحداث «المتساوية الاحتمال»؛ على سبيل المثال، إذا كنتُ أريد معرفة احتمال أن تستغرق رحلتِي الصباحية للعمل أقل من ساعة واحدة، فإنه ليس من الواضح على الإطلاق ما ينبغي أن تكون عليه الأحداث الابتدائية المتساوية الاحتمال. لا يوجد تماثُل واضح في هذا الموقف مشابه للتماثل الموجود في حالة النرد. وعلاوة على ذلك، إذا تطلبنا أن تكون الأحداث الابتدائية «متساوية الاحتمال» فسنقع في فخ التعريف الدائري؛ إذ يبدو أننا هكذا نعرِّف الاحتمال باحتمال.

ويجدر التأكيد هنا على أن كل هذه التفسيرات المختلفة للاحتمالات تتوافق مع البديهيات نفسها ويتم معالجتها بالآلية الرياضية نفسها. ما يختلف ببساطة هو طريقة رسم خريطة العالم الحقيقي؛ أي تعريف ما «يعنيه» الكائن الرياضي. أحيانًا أقول إن «الحساب» هو نفسه، ولكن «النظرية» مختلفة. وفي التطبيقات الإحصائية — كما سنرى في الفصل الخامس — يمكن للتفسيرات المختلفة أن تؤدي في بعض الأحيان إلى استخلاص استنتاجات مختلفة.

(٣) قوانين المصادفة

ذكرنا بالفعل قانونًا واحدًا من قوانين الاحتمالات؛ وهو قانون الأعداد الكبيرة. وهذا القانون يربط رياضيات الاحتمالات بالملاحظات التجريبية في العالم الحقيقي. وثمة قوانين أخرى للاحتمالات متضمنة في بديهيات الاحتمالات. وتتضمن بعض هذه القوانين المهمة للغاية مفهوم «الاستقلال».

يقال إن الحدثين مستقلان إذا كان وقوع أحدهما لا يؤثر على احتمالات وقوع الآخر؛ فحقيقة أن العملة التي قذفتُها بيدي اليسرى سوف تستقر ووجه الكتابة لأعلى بدلًا من وجه الصورة لا تؤثر على نتائج قذف العملة بيدي اليمنى، فعمليتا قذف العملة هاتان مستقلتان. وإذا كان احتمال أن العملة الموجودة في يدي اليسرى سوف تستقر ووجه الصورة لأعلى هو ١ / ٢، واحتمال أن العملة في يدي اليمنى سوف تستقر ووجه الصورة لأعلى هو ١ / ٢، فإن احتمال أن كلتا العملتين سوف تقع على وجه الصورة هو ١ / ٢ × ١ / ٢ = ١ / ٤. يسهل إدراك هذا حيث إننا نتوقع أنه في كثير من تكرارات تجربة القذف المزدوج سوف تستقر القطعة النقدية في اليد اليسرى ووجه الصورة لأعلى

فيما يقرب من نصف مرات قذفها، وأنه من بين هذه الحالات، سوف تستقر القطعة النقدية في اليد اليمنى ووجه الصورة لأعلى فيما يقرب من نصف مرات قذفها لأن نتائج عملية القذف الأولى لا تؤثر على الثانية. وبوجه عام، فإن حوالي ١ / ٤ عدد مرات القذف المزدوج من شأنه أن ينتج عنه صورة. وبالمثل، فإن حوالي ١ / ٤ عدد المرات سينتج عنه كتابة في عملة اليد اليسرى، وصورة في عملة اليد اليمنى، وحوالي ١ / ٤ عدد المرات سينتج عنه صورة في عملة اليد اليسرى، وكتابة في عملة اليد اليمنى، وحوالي ١ / ٤ عدد المرات سينتج عنه كتابة في كلتا العملتين.

في المقابل، فإن احتمال التعثر والسقوط في الشارع بالتأكيد ليس مستقلًّا عمَّا إذا كان الشارع مغطًّى بالثلوج أم لا؛ فهذان الحدثان «غير مستقلين». رأينا مثالًا آخَر للأحداث غير المستقلة في الفصل الأول؛ في حالة لسالي كلارك المأساوية التي توفي فيها طفلان في الأسرة نفسها. عندما يكون الحدثان غير مستقلين، فإننا لا نستطيع حساب احتمال وقوع كلٍّ منهما ببساطة عن طريق ضرب احتمالَي وقوعهما المنفصلين معًا. وفي الواقع، كان هذا هو الخطأ الذي كان يكمن في جوهر قضية سالي كلارك. لإدراك ذلك، دعنا نأخذ الموقف الأكثر تطرفًا لحدثين غير مستقلين تمامًا؛ وهو عندما «تحدِّد» نتائجُ أحد الحدثين «على نحو تام» نتائجَ الحدث الآخر؛ على سبيل المثال، تأمَّلْ عملية قذف عملة واحدة، والحدثان «وجه الصورة للعملة لأعلى» و«وجه الكتابة للعملة لأسفل». كلٌّ من هذين الحدثين لديه احتمال يبلغ النصف؛ فاحتمال أن العملة سوف تستقر ووجه الصورة لأعلى هو ١ / ٢، واحتمال أن العملة سوف تستقر ووجه الكتابة لأسفل هو ١ / ٢. ولكن من الواضح أنهما ليسا حدثين مستقلين. في الواقع، هما مرتبطان ارتباطًا تامًّا. فعلى أي حال، إذا كان الحدث الأول صحيحًا (الصورة لأعلى) «يجب» أن يكون الثاني صحيحًا (الكتابة لأسفل). ولأنهما مرتبطان ارتباطًا تامًّا، فإن احتمال أن يحدث كلاهما يساوي ببساطة احتمال حدوث الأول؛ وهو احتمال يبلغ النصف. وليس هذا ما نحصل عليه إذا ضربنا الاحتمالين المنفصلين البالغ كلٌّ منهما نصفًا معًا.

بصفة عامة، يعني عدم الاستقلال بين حدثين أن احتمال أن أحدهما سيحدث يعتمد على كون الآخر قد حدث أم لا.

يطلق الإحصائيون على احتمال وقوع حدثين معًا اسم «الاحتمال المشترك» لهذين الحدثين؛ على سبيل المثال، يمكننا أن نتحدث عن الاحتمال المشترك بأنني سوف أنزلق وأن الطريق مغطًّى بالثلوج. والاحتمال المشترك بين حدثين يرتبط ارتباطًا وثيقًا باحتمال

أن يقع حدثٌ ما «إذا» وقع حدثٌ آخر. هذا يسمى «الاحتمال الشرطي»؛ أي احتمال أن حدثًا ما سوف يقع نظرًا لوقوع حدث آخر. وهكذا يمكننا أن نتحدث عن الاحتمال الشرطي بأنني سوف أنزلق لأن الطريق مغطًّى بالثلوج.

إن الاحتمال (المشترك) لوقوع كلا الحدثين «أ» و«ب» هو ببساطة احتمال وقوع الحدث «أ» مضروبًا في احتمال وقوع الحدث «ب» (المشروط) نظرًا لوقوع «أ»؛ فالاحتمال (المشترك) أن الثلوج تتساقط وأنني سأنزلق هو احتمال أن الثلوج تتساقط مضروبًا في الاحتمال (المشروط) أنني سأنزلق إذا كانت الثلوج قد تساقطت.

وللتوضيح، تأمَّلْ رمية واحدة لحجر نرد وحدثين. الحدث «أ» هو أن الرقم الظاهر يقبل القسمة على ٢، والحدث «ب» هو أن الرقم الظاهر يقبل القسمة على ٣. الاحتمال المشترك لهذين الحدثين «أ» و«ب» هو احتمال أن نحصل على عدد يقبل القسمة على ٢ و٣، وهذا الاحتمال يبلغ ١ / ٦ فقط؛ إذ إن واحدًا فقط من الأرقام ١، ٢، ٣، ٤، ٥، ٦ يقبل القسمة على كل من ٢ و٣. والاحتمال المشروط للحدث «ب» نظرًا لوقوع «أ» هو احتمال الحصول على رقم يقبل القسمة على ٣ من بين الأرقام التي تقبل القسمة على ٢. حسنًا، من بين جميع الأرقام التي تقبل القسمة على ٢ (وهذا يعني، من بين ٢، ٤، ٦) رقم واحد فقط يقبل القسمة على ٣، لذلك يبلغ هذا الاحتمال الشرطي ١ / ٣. وأخيرًا، فإن احتمال الحدث «أ» هو ١ / ٢ (نصف الأرقام ١، ٢، ٣، ٤، ٥، ٦ يقبل القسمة على ٢). ومن ثَمَّ نجد أن احتمال «أ» (١ / ٢) مضروبًا في الاحتمال (الشرطي) للحدث «ب» نظرًا لوقوع «أ» (١ / ٣) هو ١ / ٦. وهو يبلغ نفس قيمة الاحتمال المشترك بالحصول على عدد يقبل القسمة على كل من ٢ و٣؛ أي الاحتمال المشترك لوقوع الحدثين «أ» و«ب».

في الواقع، الْتَقَيْنا سابقًا مفهوم الاحتمال الشرطي في الفصل الأول، في صورة مغالطة المدعي. وأشار هذا إلى أن احتمال وقوع الحدث «أ» نظرًا إلى حدوث «ب» ليس هو الاحتمال نفسه بوقوع الحدث «ب» نظرًا لوقوع «أ»؛ على سبيل المثال، احتمال أن شخصًا ما يُدِير شركة كبرى يستطيع قيادة سيارة ليس هو الاحتمال نفسه بأن الشخص الذي يستطيع قيادة سيارة يدير شركة كبرى. وهذا يقودنا إلى قانون آخَر مهمٍّ للغاية من قوانين الاحتمالات؛ وهو «مبرهنة بايز» (أو «قاعدة بايز»). تساعدنا مبرهنة بايز في ربط هذين الاحتمالين الشرطيين؛ الاحتمال الشرطي للحدث «أ» نظرًا لوقوع «ب»، والاحتمال الشرطي للحدث «ب» نظرًا لوقوع «أ».

رأينا للتوِّ أن احتمال وقوع كلا الحدثين «أ» و«ب» يساوي احتمال أن «أ» سيَقَع مضروبًا في الاحتمال (المشروط) بأن «ب» سيقع نظرًا لوقوع «أ». ولكن يمكن أيضًا

كتابة هذا على نحو معكوس؛ احتمال أن كلا الحدثين «أ» و«ب» سوف يقعان يساوي أيضًا احتمال أن «ب» سيقع مضروبًا في احتمال أن «أ» سيقع نظرًا لوقوع «ب». وتنص نظرية بايز (على الرغم من أنه عادة ما يُعبر عن ذلك على نحو مختلف) على أن هاتين الطريقتين ببساطة طريقتان بديلتان لكتابة الاحتمال المشترك للحدثين «أ» و«ب»؛ أي إن احتمال «أ» مضروبًا في احتمال «ب» نظرًا لوقوع الحدث «أ» يساوي احتمال «ب» مضروبًا في احتمال «أ» نظرًا لوقوع الحدث «ب». وكلاهما يساوي الاحتمال المشترك بين «أ» و«ب». في مثال «رئيس الشركة الذي يقود سيارة»، تكافئ نظرية بايز قول إن احتمال إدارتك لشركة كبرى نظرًا إلى أنك تستطيع قيادة سيارة، مضروبًا في احتمال أن تتمكن من قيادة سيارة، يساوي احتمال أنك تستطيع قيادة سيارة نظرًا إلى أنك رئيس شركة، مضروبًا في احتمال كونك رئيس شركة. وكلاهما يساوي الاحتمال المشترك لكونك رئيس شركة وقادرًا على قيادة سيارة.

ينص قانون آخر للاحتمالات على أنه إذا كان يمكن وقوع أحد الحدثين، ولكن لا يمكن أن يقع كلاهما معًا، فإن احتمال أن أحدهما سيقع هو مجموع الاحتمالين المنفصلين لوقوع كلٍّ منهما. إذا قذفتَ عملة — ومن المؤكد أنها لا يمكن أن تظهر وجه الكتابة والصورة في الوقت ذاته — فإن احتمال ظهور وجه الصورة «أو» وجه الكتابة هو مجموع احتمال أن وجه الصورة سوف يظهر واحتمال أن وجه الكتابة سوف يظهر. إذا كانت العملة متزنة، فإن كلا هذين الاحتمالين المنفصلين هو النصف، وهكذا فإن الاحتمال الكلي لظهور وجه الصورة ووجه الكتابة هو ١. هذا الأمر يبدو معقولًا تمامًا؛ إذ يتوافق الرقم ١ مع اليقين، ومن المؤكد أنه يجب أن يَظهَر وجه الصورة أو وجه الكتابة (أفترض أنه لا يمكن أن ينتهي الأمر بوقوف العملة على حافتها!) وبالعودة إلى مثال رمي النرد: كان احتمال الحصول على عدد زوجي هو مجموع احتمالات الحصول على أيٍّ من الأرقام ٢ أو ٤ أو ٦؛ لأنه لا يمكن أن يقع أي من هذه الأحداث معًا (ولا توجد أي طرق أخرى للحصول على عدد زوجي برمية واحدة للنرد).

(٤) المتغيرات العشوائية وتوزيعاتها

رأينا في الفصل الثاني كيف يمكن استخدام الملخصات الإحصائية البسيطة لاستخراج المعلومات من مجموعة كبيرة من قِيَم متغيِّرٍ ما، بحيث تكثف المجموعة ليكون توزيع القِيَم سهلَ الفهم. إن أي مجموعة بيانات حقيقية تكون محدودة في الحجم؛ فلا يمكن

أن تحتوي إلا على عدد محدود من القِيَم. هذه المجموعة المحدودة قد تمثل قيم كافة الأشياء من النوع الذي نُخضعه للدراسة (مثل درجات جميع لاعبي دوري كرة القدم في سنة معينة) أو قد تمثل قيم بعض الأشياء فحسب؛ أيْ إنها «عينة». ورأينا أمثلة على هذا عندما تناولنا مسح العينات.

العينة هي مجموعة فرعية من «مجموعة القيم» الكاملة الخاضعة للدراسة. في بعض الحالات، تكون المجموعة الكاملة غيرَ واضحة التعريف، وربما تكون ضخمة أو حتى لا نهائية؛ لذلك لا يكون لدينا خيار اللجوء إلى عينة؛ على سبيل المثال، في تجارب قياس سرعة الضوء، في كل مرة آخذ فيها القياس أتوقع الحصول على قيمة مختلفة قليلًا؛ وذلك ببساطة بسبب عدم الدقة في عملية القياس. ويمكنني — على الأقل من حيث المبدأ — المضي قدمًا في أخذ القياسات إلى الأبد؛ وهذا يعني أن مجموعة القياسات المحتملة الكاملة لا نهائية. وبما أن هذا أمر مستحيل، يجب أن أرضى بعينة محدودة من القياسات. وسوف تُستخرَج هذه القياسات من المجموعة الكاملة للقيم التي يحتمل أن أحصل عليها. وفي حالات أخرى، تكون المجموعة الكاملة محدودة؛ على سبيل المثال، في دراسة للسِّمْنة بين الذكور في بلدة معينة، تكون مجموعة الخاضعين للدراسة محدودة، ورغم أنني من حيث المبدأ أستطيع وزن كل واحد منهم في المدينة، ففي الممارسة العملية ربما لن أريد ذلك، وسوف أستخدم عينة. ومرة أخرى، كل قيمة في عينتي مأخوذة من المجموعة الكاملة للقيم الممكنة.

في كلٍّ من هذه الأمثلة، كل ما أعرفه قبل أن آخذ كل قياس هو أنه سيكون له قيمةٌ ما من مجموعة القيم الكاملة الممكنة. ستحدث كل قيمة باحتمال معين، ولكني لا أستطيع أن أحدده أكثر من ذلك، وربما لا أعرف ما هو هذا الاحتمال. وبالتأكيد لا أستطيع أن أحدد بالضبط القيمة التي سوف أحصل عليها في القياس التالي لسرعة الضوء أو ماذا سيكون وزن الرجل التالي الذي سأقيسه. وبالمثل، في رمي النرد، أعلم أن النتيجة يمكن أن تكون ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦، وهنا أعرف أن هذه الاحتمالات متساوية (فنردي مكعب مثالي)، ولكن بخلاف ذلك لا أستطيع أن أحدد العدد الذي سيَظهَر. وعلى غرار قياسات السرعة والوزن، فإن النتيجة عشوائية؛ ولهذا السبب تُسمَّى هذه المتغيرات «متغيرات عشوائية».

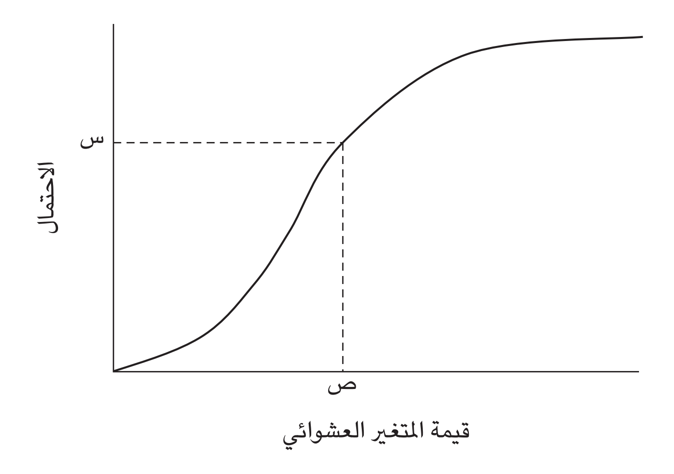
الْتَقَيْنا من قبلُ بالفعل مفهومَ المقاييس التجزيئية؛ على سبيل المثال، في حالة المُؤَيَّات، المُؤَيُّ ٢٠ من التوزيع هو القيمة التي يكون ٢٠٪ من قيم البيانات أصغر

منها، والمُؤَي ٨ هو القيمة التي يكون ٨٪ من قيم البيانات أصغر منها، وهكذا. وعمومًا، المُؤَيُّ K يكون $k\%$ من العينة أصغر منه. ويمكننا تصور تحديد مُؤَيَّات مماثلة، ليس للعينة التي نرصدها فحسب، ولكن بالنسبة للمجموعة الكاملة من القيم التي يمكننا رصدها. إذا عرفنا المُؤَيَّ ٢٠ للمجموعة الكاملة من القيم، حينها فسنعرف أن القيمة المأخوذة عشوائيًّا من المجموعة الكاملة لديها احتمال ٠٫٢٠ أن تكون أصغر من هذا المُؤَيِّ. وعمومًا، إذا عرفنا كل مُؤَيَّات المجموعة الكاملة، فسنعرف احتمال استخراج قيمة في آخَر ١٠٪ أو ٢٥٪ أو ١٦٪ أو ٩٨٪ أو أي نسبة مئوية أخرى نهتمُّ باختيارها؛ وهذا يعني أننا حينها نعرف كل شيء عن توزيع القيم الممكنة التي نستطيع الحصول عليها. لن نعرف ما القيمة التالية التي سنحصل عليها، ولكن سنعرف احتمال أنها ستكون في أصغر ١٪ من القيم في المجموعة الكاملة، أو في أصغر ٢٪، وما شابه ذلك.

يوجد اسم لمجموعة مُؤَيَّات التوزيع الكاملة؛ إذ يطلق عليها اسم «توزيع الاحتمال التراكمي»، وهو «توزيع احتمال» لأنه يخبرنا «باحتمال» الحصول على قيمة أقل من أي قيمة نختارها، وهو «تراكمي» لأنه من الواضح أن احتمال الحصول على قيمة أقل من القيمة «س» يزداد كلما زادت «س». في مثال أوزان الذكور، لو كنتُ أعرف أن احتمال اختيار رجل وزنه أقل من ٧٠ كيلوجرامًا هو ١ / ٢، فإنني حينها سأعلم أن احتمال اختيار رجل وزنه أقل من ٨٠ كيلوجرامًا هو أكثر من ١ / ٢ لأنه يمكنني أن أختار من بين كل أولئك الذين يقلُّ وزنهم عن ٧٠ كيلوجرامًا، وكذلك أولئك الذين يكون وزنهم بين ٧٠ كيلوجرامًا و٨٠ كيلوجرامًا. وعند الحد الأقصى، فإن احتمال الحصول على قيمة أقل من أو تساوي أكبر قيمة في مجموعة القيم الكاملة هو ١؛ أي إنه حدث مؤكد.

تتضح هذه الفكرة في الشكل ٤-١؛ ففي هذا الشكل، تُمثَّل قيم المتغير العشوائي (فكر في الوزن) على المحور الأفقي، ويُمثَّل احتمال القيم الأصغر على المحور الرأسي. ويبين المنحنى احتمال أن تكون القيمة المختارة عشوائيًّا — بالنسبة لأي قيمة معينة للمتغير العشوائي — أصغر من هذه القيمة المعينة.

يخبرنا توزيع الاحتمال التراكمي لمتغير عشوائي باحتمال أن تكون قيمة مختارة عشوائيًّا «أقل» من أي قيمة أخرى. والطريقة البديلة للنظر إلى الأمور هي أن ننظر إلى احتمال أن قيمة مختارة عشوائيًّا سوف تَقَع «بيْنَ» أي قيمتين معينتين. وتتمثل هذه الاحتمالات على نحو ملائم في سياق النطاق الواقع بين قيمتين تحت منحنى «الكثافة» الاحتمالية؛ على سبيل المثال، يبين شكل ٤-٢ منحنى «الكثافة الاحتمالية»، ويبين منطقةً

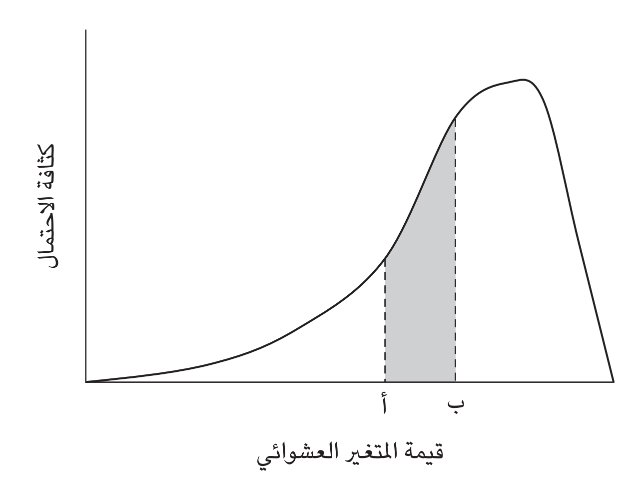


شكل ٤-١: توزيع الاحتمال التراكمي.

(مظللة) تحت المنحنى بين نقطتي «أ» و«ب» ممثلة احتمال أن قيمة مختارة عشوائيًّا سوف تقع بين «أ» و«ب»؛ وعلى سبيل المثال، عند استخدام مثل هذا المنحنى لتوزيع أوزان الرجال في بلدتنا، يمكن أن نجد احتمال أن يقع رجل مختار عشوائيًّا بين ٧٠ كيلوجرامًا و٨٠ كيلوجرامًا، أو أي زوج آخر من القيم، أو فوق أي قيمة نريدها أو تحتها. وعلى نحو عام، من المرجح أن تحدث القيم المختارة عشوائيًّا في المناطق التي تكون فيها الاحتمالية أكثر كثافة؛ أي حيث يكون منحنى الكثافة الاحتمالية في أعلاه.

لاحظ أن المساحة الكلية تحت المنحنى في شكل ٤-٢ يجب أن تكون ١ — المتوافق مع اليقين — ويجب أن تكون للقيمة المختارة عشوائيًّا قيمة «جزئية» منها.

تمتلك منحنيات التوزيع للمتغيرات العشوائية أشكالًا مختلفة؛ فاحتمال أن امرأة مختارة عشوائيًّا سوف يكون وزنها بين ٧٠ كيلوجرامًا و٨٠ كيلوجرامًا عادة لا يكون هو نفسه احتمال أن رجلًا مختارًا عشوائيًّا سيكون وزنه بين هاتين القيمتين. وربما نتوقَّع أن منحنى توزيع أوزان النساء سيأخذ قِيَمًا كبيرة في الأوزان الأصغر مما هي الحال بالنسبة لمنحنى الرجال.



شكل ٤-٢: دالة الكثافة الاحتمالية.

تمتلك بعض الأشكال أهمية خاصة، وتوجد أسباب عديدة لذلك؛ ففي بعض الحالات، تظهر أشكال معينة، أو أشكال مقاربة للغاية لهذه الأشكال، على نحو طبيعي. بينما في حالات أخرى، تنشأ التوزيعات كنتائج لقوانين الاحتمالات.

لعل أبسط التوزيعات هو «توزيع برنولي». وهذا التوزيع يمكن أن يتخذ قيمتين فحسب، قيمة لها احتمال p، مثلًا، والأخرى لها احتمال $1-p$. وبما أنه لا يمكن أن يتخذ إلَّا قيمتين فقط، فمن «المؤكد» أن إحدى القيمتين سوف تظهر؛ ومن ثَمَّ فإن مجموع احتمالات هاتين النتيجتين يساوي ١. لدينا بالفعل أمثلة أوضحت لماذا يُعَدُّ هذا التوزيع مفيدًا؛ فالحالات التي لا ينتج عنها إلا نتيجتان شائعة جدًّا؛ مثل قذف العملة التي ينتج عنها إمَّا وجه الصورة وإمَّا وجْه الكتابة، وعملية الولادة التي تكون نتيجتها إما ذكرًا وإما أنثى. في هاتين الحالتين، تمتلك p قيمة ١ / ٢ أو ما يقرب من ١ / ٢. ولكن يوجد عدد كبير من الحالات الأخرى التي لا يوجد لها سوى نتيجتين محتملتين: نعم/لا، جيد/سيئ، افتراضي أو غير افتراضي، انكسار أو عدم انكسار، توقف/حركة، وما شابه ذلك.

يوسع «التوزيع ذو الحدين» توزيع برنولي؛ فإذا قذفنا عملة ثلاث مرات، ربما يظهر وجه الصورة مرة أو مرتين أو ثلاث مرات أو لا يظهر أبدًا. وإذا كان لدينا ثلاثة موظفين

في مركز اتصالات يجيبون على نحو مستقلٍّ على المكالمات عندما تَرِدُ، فإنه من الممكن أن يكون واحد أو اثنان أو الثلاثة مشغولين أو لا يكون أحدهم مشغولًا في أي لحظة معينة. يخبرنا التوزيع ذو الحدين باحتمال حصولنا على كل رقم من هذه الأرقام ٠، أو ١، أو ٢، أو ٣. وبطبيعة الحال، فإنه يطبق على نحو عام، وليس فقط على المجموع الكلي لثلاثة أحداث. فإذا قذفنا عملة مائة مرة، فإن التوزيع ذا الحدَّيْن يخبرنا أيضًا باحتمالات أننا سنحصل على كلٍّ من ٠، ١، ٢، ...، ١٠٠ وجه صورة.

تصل رسائل البريد الإلكتروني إلى جهاز الكمبيوتر الخاص بي عشوائيًّا. وتصل خلال العمل الصباحي — في المتوسط — (مثلًا) بمعدل خمس رسائل في الساعة، ولكن عدد الرسائل التي تصل في كل ساعة يمكن أن ينحرف عن هذا المعدل على نحو كبير جدًّا؛ إذ يصل في بعض الأحيان عشر رسائل، وفي أحيان أخرى لا تصل أي رسالة. يمكن استخدام «توزيع بواسون» لوصف التوزيع الاحتمالي لعدد رسائل البريد الإلكتروني التي تصل في كل ساعة. ويمكن أن يخبرنا باحتمال (إذا كانت رسائل البريد الإلكتروني تصل على نحو مستقل وكان المعدل العام لوصولها ثابتًا) عدم وصول أي رسالة، أو وصول رسالة واحدة، أو رسالتين، وما إلى ذلك. وهذا التوزيع يختلف عن التوزيع ذي الحدين؛ لأنه على الأقل من حيث المبدأ لا يوجد حد أعلى للعدد الذي يمكن أن يصل في أي ساعة. ففي حالة قذف العملة مائة مرة، لا يمكننا رؤية أكثر من ١٠٠ وجه صورة، ولكن يمكن أن يصلني (في يوم سيئ للغاية!) أكثر من ١٠٠ رسالة بريد إلكتروني في ساعة واحدة.

حتى الآن، كل التوزيعات الاحتمالية التي ذكرتُها هي لمتغيرات عشوائية «منفصلة» (أو متقطعة)؛ أي إن المتغيرات العشوائية لا تأخذ سوى قيمٍ معينة (قيمتين في حالة توزيع برنولي، عدد من القيم يعتمد على عدد مرات قذف العملة/عدد المشغلين في حالة التوزيع ذي الحدين، والأعداد الصحيحة ٠، ١، ٢، ٣، ... في حالة توزيع بواسون). ثمة متغيرات عشوائية أخرى «متصلة» (أو مستمرة)، ويمكن أن تأخذ أي قيمة من النطاق؛ فعلى سبيل المثال، الطول يمكن أن يأخذ أي قيمة داخل نطاق معين (رهنًا بدقة أداة القياس)، ولا يقتصر، مثلًا، على ٤ أو ٥ أو ٦ أقدام.

إذا كان المتغير العشوائي يمكن أن يأخذ قيمًا ضمن فترة محدودة فقط (على سبيل المثال بين ٠ و١) وإذا كان «من المحتمل على نحو متساوٍ» أن يأخذ أي قيمة من القيم في تلك الفترة، يقال إنه يتبع «توزيعًا منتظمًا»؛ على سبيل المثال، إذا كان ساعي البريد يصل دائمًا في الفترة من ١٠ صباحًا حتى ١١ صباحًا، ولكن بطريقة لا يمكن التنبؤ بها

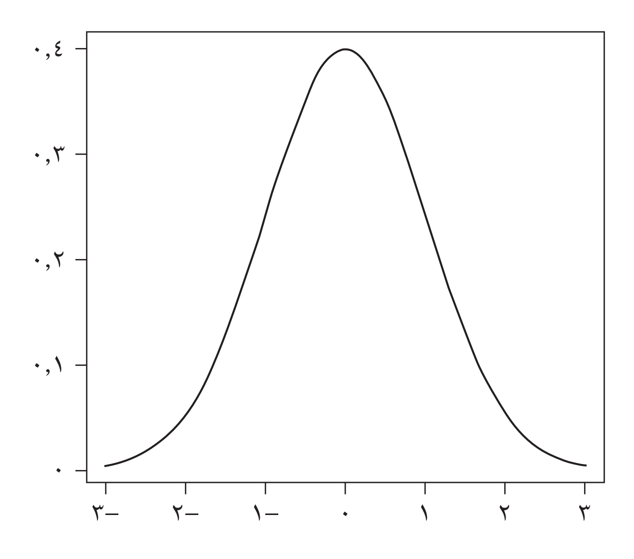
تمامًا (فمن المحتمل أن يصل في الفترة بين ١٠:٠٥ حتى ١٠:١٠ بالقدر نفسه لاحتمال وصوله في أي فترة خمس دقائق أخرى على سبيل المثال)، فإن توزيع وقت وصوله خلال هذه الفترة يكون منتظمًا.

يمكن لبعض المتغيرات العشوائية أن تأخذ أي قيمة موجبة؛ شأن المدة الزمنية لظاهرةٍ ما على سبيل المثال. لتوضيح ذلك، تأمَّلِ الفترة الزمنية التي تَظَلُّ فيها المزهريات الزجاجية سليمة قبل أن تنكسر. المزهريات الزجاجية لا تتقدم في العمر، لذلك ليس من المرجح بدرجة أكبر أن تنكسر المزهرية المفضَّلة لديك في العام المقبل إذا كان عمرها ٨٠ سنة، من أن تنكسر في العام المقبل إذا كان عمرها ١٠ سنوات فقط (في حالةِ تَساوي كل العوامل الأخرى). قارِنْ ذلك مع احتمال وفاة شخص يبلغ من العمر ٨٠ سنة في العام المقبل في مقابل احتمال وفاة شخص يبلغ من العمر ١٠ سنوات في العام المقبل. بالنسبة لمزهرية زجاجية، إذا لم تتحطَّم في الفترة الزمنية t، فإن احتمال أن تتحطم في اللحظة التالية هو الاحتمال نفسه، مهما كانت قيمة t (مرة أخرى، كل العوامل الأخرى متساوية). يقال هنا إن عمر المزهريات الزجاجية يتبع «توزيعًا أُسِّيًّا». في الواقع، توجد أعداد هائلة من تطبيقات التوزيعات الأُسِّيَّة، وليس أعمار المزهريات الزجاجية فحسب!

ولعل الأكثر شهرة بين التوزيعات المستمرة هو «التوزيع الطبيعي» أو «توزيع جاوس». وغالبًا ما يوصف على نحو عام في سياق شكله العام: «شكل الجرس»، كما هو مبين في الشكل ٤-٣. وهذا يعني أن احتمال حدوث القيم الموجودة في الوسط أكبر من احتمال حدوث القيم في الطرفين البعيدين عن الوسط. يوفر التوزيع الطبيعي تقريبًا جيدًا لكثير من التوزيعات التي تحدث طبيعيًّا؛ على سبيل المثال، توزيع أطوال عينة عشوائية من الرجال البالغين يتبع توزيعًا طبيعيًّا تقريبًا.

يظهر التوزيع الطبيعي أيضًا في كثير من الأحيان بمَظهَر النموذج الجيد لشكل توزيع إحصائيات العينة (مثل الملخصات الإحصائية المذكورة في الفصل الثاني) عندما تنطوي على عينات كبيرة. على سبيل المثال، لنفترض أننا أخذنا على نحو متكرر عينات عشوائية من توزيعٍ ما، وحسبنا متوسط كل عينة من هذه العينات. بما أن كل عينة مختلفة، فإننا نتوقع أن يكون كل متوسط مختلفًا؛ أي سيكون لدينا توزيع للمتوسطات. وإذا كانت كل عينة كبيرة بما فيه الكفاية، فسيتضح أن هذا التوزيع للمتوسطات هو توزيع طبيعي تقريبًا.

أشرتُ في الفصل الثاني إلى أن الإحصاء ليس مجرد مجموعة من الأدوات المعزولة، ولكنه لغة متصلة. وتنطبق نقطة مماثلة على التوزيعات الاحتمالية. فعلى الرغم من أنني



شكل ٤-٣: التوزيع الطبيعي.

ذكرتُها كلًّا على حِدَة آنِفًا، فإن الحقيقة هي أن توزيع برنولي يمكن اعتباره حالة خاصة من التوزيع ذي الحدَّيْن (فهو توزيع ذو حدَّيْن عندما لا يوجد سوى نتيجتين محتملتين فحسب). وبالمثل، على الرغم من أن العمليات الرياضية التي تُظهِر هذا تتخطَّى حجم هذا الكتاب، فإن توزيع بواسون يمثِّل حالة متطرِّفة من التوزيع ذي الحدَّيْن، ويشكِّل توزيع بواسون والتوزيع الأُسِّي زوجًا طبيعيًّا، ويصبح التوزيع ذو الحدين أكثر وأكثر شبهًا بالتوزيع الطبيعي كلما زاد الحد الأقصى لعدد الأحداث، وهكذا. وهذه التوزيعات في حقيقتها جزء من وحدة رياضية كاملة متكاملة.

لقد وصفتُ التوزيعات السابقة بالقول إن لها أشكالًا مختلفة. وفي الواقع، يمكن وصف هذه الأشكال على نحو ملائم. فرأَيْنا أن توزيع برنولي يتميز بوجود القيمة p. وهذا يُخبِرنا باحتمال أننا سوف نحصل على نتيجة معينة. وتتوافق قيم p المختلفة مع توزيعات برنولي مختلفة؛ فيمكننا صياغة نتائج قذف عملة عن طريق توزيع برنولي مع احتمال ظهور وجه الصورة — p — مساويًا النصف، وصياغة احتمال وقوع حادث

لسيارة في رحلة واحدة بواسطة توزيع برنولي مع p تساوي قيمة صغيرة جدًّا (كما آمل!). وفي مثل هذه الحالة، تسمى p «مَعْلَمة» (أو بارامترًا).

وتتميز التوزيعات الأخرى أيضًا بوجود معلمات تؤدي الدور نفسه؛ إذ تعرفنا بالضبط بعضو عائلة التوزيعات الذي نتحدث عنه. لنرى كيفية ذلك، دَعْنا نَعُدْ خطوة إلى الوراء ونتذكر قانون الأعداد الكبيرة. ينص هذا القانون على أننا إذا قمنا بملاحظات مستقلَّة متكررة لحدث له نتيجة A باحتمال p ونتيجة B باحتمال $1 - p$، فإننا يجب أن نتوقع أن نسبة مرات ملاحظة النتيجة A تقترب أكثر وأكثر من p كلما زاد عدد الملاحظات التي نقوم بها. تُعمَّم هذه السمة بطرق مهمة. فعلى وجه الخصوص، لنفترض أنه بدلًا من ملاحظة حدث له نتيجتان محتملتان فحسب، لاحظنا حدثًا يمكن أن يأخذ أي قيمة من توزيع على مجموعة من القيم؛ على سبيل المثال، ربما يأخذ أي قيمة في الفترة [٠، ١]. ولنفترض أننا أخذنا مجموعات من القياسات n من مثل هذا التوزيع على نحو متكرر. يخبرنا قانون الأعداد الكبيرة أيضًا أنه ينبغي لنا أن نتوقع أن يقترب متوسط القياسات n من قيمة ثابتة كلما كانت n أكبر. وفي الواقع، يمكننا تصور زيادة n دون حد، وفي هذه الحالة من المنطقي أن نتحدث عن متوسط عينة غير محدودة مستمدة من التوزيع؛ بل وحتى متوسط التوزيع نفسه. فعلى سبيل المثال، باستخدام هذه الفكرة يمكن أن نتحدث عن متوسط التوزيع الأُسِّي نفسه وليس عن متوسط «عينة مأخوذة من التوزيع الأسي» فحسب. وتمامًا كما ستمتلك توزيعات برنولي المختلفة معلمات p مختلفة، فإن التوزيعات الأسية المختلفة سوف تمتلك متوسطات مختلفة. وحينها يكون المتوسط معلمة للتوزيع الأسي.

رأينا في مثال سابق أن التوزيع الأسي كان نموذجًا معقولًا «لعمر» المزهريات الزجاجية (تحت ظروف معينة)، والآن يمكننا أن نتصور أن لدينا مجموعتَين من هذه المزهريات؛ مجموعة تتكون من مزهريات صلبة مصنوعة من زجاج سميك للغاية، ومجموعة ثانية تتكون من مزهريات هَشَّة مصنوعة من زجاج رقيق للغاية. من الواضح أنه في المتوسط، مزهريات المجموعة الأولى من المرجح أن تعيش لفترة أطول من مزهريات المجموعة الثانية. كل مجموعة من المجموعتَين لها معلمة مختلفة.

يمكننا تحديد المعلمات الخاصة بالتوزيعات الأخرى على نحو مشابِهٍ؛ فنتصور حساب ملخصات إحصائية لعينات بحجم لا نهائي مستمدة من التوزيعات؛ على سبيل المثال، يمكننا أن نتصور حساب متوسطات عينات كبيرة لا نهائية مستمدة من أعضاء

الأسرة العادية للتوزيعات. إلا أن الأمور أكثر تعقيدًا قليلًا هنا؛ لأن أعضاء هذه الأسرة من التوزيعات لا تتحدَّد على نحو فريد بواسطة معلمة واحدة؛ فهي تتطلب معلمتين. في الواقع، المتوسط والانحراف المعياري للتوزيعات سيكونان كافيَيْنِ؛ إذ سيعملان معًا على تحديد أي أعضاء العائلة نتحدث عنه على نحو فريد.

نُقِّح قانون الأعداد الكبيرة أكثر من ذلك. تخيَّلِ استخراج العديد من مجموعات القيم من توزيعٍ ما، بحيث تكون كل مجموعة بالحجم n، واحسب المتوسط لكل مجموعة. حينها ستكون المتوسطات نفسها عينة من التوزيع؛ توزيع القيم المحتملة لمتوسط عينة بالحجم n. تخبرنا «مبرهنة النهاية المركزية» أن توزيع هذه المتوسطات نفسها يتبع تقريبًا توزيعًا طبيعيًّا، وهذا التقريب يزداد أكثر وأكثر كلما زادت قيمة n. في الواقع، إنها تُخبِرنا أكثر من هذا؛ إذ تخبرنا أيضًا أن متوسط توزيع المتوسطات هذا يتطابق مع متوسط المجموعة الكاملة للقيم، وأن التباين في توزيع المتوسطات يساوي فقط $1/n$ ضعف حجم تباين توزيع المجموعة الكاملة. ويتضح أن هذا مفيد للغاية في الإحصاء؛ لأنه يعني أننا يمكننا تقدير متوسط المجموعة الكاملة بأكبر قدر من الدقة نرغب فيه فقط عن طريق أخذ عينة كبيرة بما يكفي (أخذ n كبيرة بما فيه الكفاية)؛ حيث تخبرنا مبرهنة النهاية المركزية مدى حجم العينة الذي يجب أن نصل إليه لتحقيق احتمال كبير للوصول لهذه الدقة. وبشكل أعم، يُعَدُّ المبدأ القائل بأننا نستطيع الحصول على تقديرات أفضل وأفضل من خلال أخذ عينات أكبر مبدأً قويًّا للغاية. وقد رأينا بالفعل إحدى الطرق التي تُستخدم فيها هذه الفكرة على نحو عملي حين تناولنا موضوع مسح العينات في الفصل الثالث.

إليك مثالًا آخَر. في علم الفلك، تكون الأجرام السماوية البعيدة خافتة جدًّا، وتكون المشاهدات معقَّدة بسبب التقلبات العشوائية في الإشارات. ومع ذلك، إذا أخذنا العديد من الصور للجرم نفسه وراكَبْناها بعضها فوق بعض، فإن الأمر يُشبِه حساب متوسط العديد من القياسات للشيء نفسه، وكل قياس مستمد من التوزيع نفسه ولكن بوجود مكوِّن عشوائي إضافي. وباستخدام قوانين الاحتمالات المذكورة سابقًا يتم التخلص من العشوائية، وتبقى رؤية واضحة للإشارة الأساسية؛ أي الجرم السماوي.

الفصل الخامس

التقدير والاستدلال

الإحصاء هو الفلسفة التطبيقية للعلوم.

إيه بي ديفيد

رأينا في الفصل الأول أن الإحصائيات تلعب دورًا مزدوجًا يتمثل في تلخيص البيانات واستخراج الاستنتاجات من البيانات، واستكشفنا بعض الأدوات البسيطة لتلخيص البيانات في الفصل الثاني. وفي هذا الفصل، وباستخدام مفاهيم الاحتمالات المذكورة في الفصل الرابع، سوف نتناول التقدير والاستدلال؛ أي، سندرس طرق تحديد قيمة الكميات التي لا يمكننا ملاحظتها بالفعل، وتقديم إفادات عنها. إليك بعض الأمثلة:

مثال ١: لتحديد سرعة الضوء، سنقوم بتنفيذ بعض طرق القياس. لا توجد طريقة قياس مثالية، وإذا كررنا هذه العملية فربما سنحصل على قيمة مختلفة قليلًا. وتكرار القياس مائة مرة من المرجح أن يعطينا مائة قيمة مختلفة قليلًا. وهدفُنا إذن هو استخدام هذه العينة من القيم لتقدير سرعة الضوء الحقيقية، دون أن يشوبها شائبة من خطأ القياس.

مثال ٢: في تجربة سريرية عشوائية بسيطة، ربما نُعطِي دواءً جديدًا لعيِّنة من المرضى ودواءً معياريًّا لعينة أخرى. وبناءً على ملاحظات الآثار لدى هاتين المجموعتَين من المرضى سوف نرغب في تقديم إفادة، أو استدلال، حول الفعالية النسبية للدواء الجديد. بعبارة أخرى، نرغب في تقدير مدى كبر الفارق في فعالية الدواءين الذي قد نتوقَّعه إذا كنَّا قد وصفْنا كل دواء من الدواءين للمجموعة الكاملة من المرضى الخاضعين

للدراسة. وسنرغب أيضًا على نحو مثالي في الحصول على بعض المؤشرات حول مدى ثقتنا في حجم التقدير.

مثال ٣: في دراسة للبطالة في لندن، ستكون مقابلة الجميع غير قابلة للتطبيق؛ لذلك ستُجرى مقابلات مع عينة من الأشخاص، بهدف استخدام ردود هذه العينة لتقديم إفادة عامة عن لندن بأكملها؛ أي إننا نود تقدير البطالة في لندن بأكملها باستخدام بيانات العينة.

مثال ٤: على نحو أكثر تجريدية، قدمتُ في الفصل الرابع فكرة «مَعْلَمة» التوزيع. وشاهدنا مثال عائلة برنولي من التوزيعات؛ حيث يستطيع متغير عشوائي أن يأخذ القيمة ٠ أو ١، وحيث كانت p معلمة تعطي احتمال ملاحظة القيمة ١. كما رأينا أيضًا مثالًا على التوزيع الطبيعي، والذي كان يمتلك معلمتين؛ هما الانحراف المعياري والمتوسط. وربما يكون هدفنا هو تقدير قيمة هذه المعلمة؛ على سبيل المثال، ربما تدرس عالِمة أنثروبولوجيا أطوال مجموعة معينة من الأشخاص، وربما تكون مستعِدَّة لافتراض أن الأطوال مُوزَّعة طبيعيًّا، ولكن لتوصيف التوزيع توصيفًا تامًّا سوف تحتاج إلى معرفة المتوسط والانحراف المعياري لهذا التوزيع. وربما ترغب في استخدام أطوال عينة أشخاص من المجموعة لتقدير المتوسط والانحراف المعياري للمجموعة بأكملها.

(١) تقدير النقطة

عَرَضَ عليَّ صديق لي الصفقةَ التالية: سوف يقذف عملة معدنية على نحو متكرر، وكلما ظهر وجه الصورة سوف يعطيني ١٠ جنيهات استرلينية، ولكن كلما ظهر وجه الكتابة سوف أعطيه ٥ جنيهات استرلينية.

يبدو هذا للوهلة الأولى وكأنه صفقة جيدة بالنسبة لي. فرغم كل شيء، من المعروف جيدًا أنه من المرجح على نحو متساوٍ أن تستقر العملة ووجه الصورة أو الكتابة لأعلى (احتمال ظهور وجه الصورة يساوي ١ / ٢)؛ لذلك من المحتمل أن أفوز بعشرة جنيهات استرلينية بقدر احتمال خسارة خمسة جنيهات استرلينية في كل قذفة للعملة. وفي المتوسط، سوف أكون فائزًا.

ولكن بعد ذلك ساورتْني الشكوك. لماذا يقدم لي صفقة يبدو أنها في صالحي للغاية؟ بدأتُ أشكُّ في أنه ربما عبث بالعملة؛ بحيث يكون احتمال ظهور وجه الصورة في الواقع

أقلَّ من النصف. فعلى أي حال، إذا كان احتمال ظهور وجه الصورة في الحقيقة ضئيلًا للغاية، بحيث إنه نادرًا ما يَظهَر، يمكن أن تكون الصفقة سيئة بالنسبة لي. لمعرفة هذا، سأرغب في تقدير هذا الاحتمال. عرض صديقي — الكريم للغاية ولكنه لا يعرف شيئًا عن الإحصاء — قذْف العملة ستَّ مرات، حتى أستطيع أن أرى الوجه الذي سيَظهَر في كل مرة. وهدفي إذن هو استخدام هذه البيانات لتقدير احتمال أن العملة ستستقر ووجه الصورة لأعلى في عمليات القذف المستقبلية.

لنفترض أن العملة خضعت للتلاعب، وأن احتمال ظهور وجه العملة في أي قذفة واحدة كان ١ / ٣ فقط. وبما أن قذفات العملة مستقلة بعضها عن بعض (نتيجة القذفة الواحدة لا تؤثر على نتيجة أي قذفة أخرى)، فإننا نعلم أن احتمال ظهور وجه الصورة في قذفتين هو ببساطة ناتج ضرب احتمال ظهور وجه الصورة في كلتا المرتين: ١ / ٣ × ١ / ٣ = ١ / ٩. وبالمثل، بما أن احتمال ظهور وجه الكتابة هو ١ − ١ / ٣ = ٢ / ٣، فإن احتمال ظهور وجه الصورة متبوعًا بظهور وجه الكتابة سيكون حاصل ضرب ١ / ٣ و٢ / ٣؛ وهو ٢ / ٩. وعمومًا، بافتراض أن احتمال ظهور وجه الصورة في كل قذفة هو ١ / ٣، يمكننا حساب احتمال الحصول على أي تسلسل لوجهَي الصورة والكتابة؛ وعلى وجه الخصوص، تسلسُل مماثل لذلك الذي يظهر في القذفات الست التي رأيناها بالفعل؛ على سبيل المثال، إذا أظهرت القذفات الست التسلسلَ ص – ك – ص – ك – ك – ك، فإن احتمال الحصول على تسلسل متطابق بالمصادفة سيكون ١ / ٣ × ٢ / ٣ × ١ / ٣ × ٢ / ٣ × ٢ / ٣ × ٢ / ٣ = ١٦ / ٧٢٩؛ أي تقريبًا ٠,٠٢٢.

يمكننا بالطريقة نفسها حساب احتمال الحصول على تسلسل ص – ك – ص – ك – ك – ك إذا كان احتمال ظهور وجه الصورة في كل قذفة يساوي فعليًّا أي قيمة أخرى؛ على سبيل المثال، إذا كان احتمال ظهور وجه الصورة ١ / ٢ (ومن ثم فإن احتمال ظهور وجه الكتابة يساوي ١ − ١ / ٢ = ١ / ٢)، فإن احتمال الحصول على مثل هذا التسلسل هو ١ / ٢ × ١ / ٢ × ١ / ٢ × ١ / ٢ × ١ / ٢ × ١ / ٢ = ١ / ٦٤؛ أي تقريبًا ٠,٠١٦، وإذا كان احتمال ظهور وجه الصورة هو ١ / ١٠، فإن احتمال الحصول على مثل هذا التسلسل يقرب من ٠,٠٠٧ وهكذا.

هدفنا الآن هو تقدير احتمال ظهور وجه الصورة في أي قذفة مستقبلية؛ أي إننا نرغب في اختيار قيمة واحدة — ١ / ٣ أو ١ / ٢ أو ١ / ١٠ أو أيًّا ما تكون — كتقدير لهذا الاحتمال. وعند النظر إلى الحسابات السابقة، نرى أن احتمال الحصول على النتائج

المرصودة لست قذفات هو ٠٫٠٢٢ إذا كان الاحتمال الحقيقي لظهور وجه الصورة هو ١ / ٣، في حين أنه لا يتجاوز ٠٫٠١٦ إذا كان الاحتمال الحقيقي لظهور وجه الصورة هو ١ / ٢، وهو أقل من ذلك — ٠٫٠٠٧ فقط — إذا كان الاحتمال الحقيقي لظهور وجه الصورة هو ١ / ١٠. ما يعنيه هذا هو أنه من الأكثر ترجيحًا أن نحصل على نتائج القذفات الست المرصودة إذا كان الاحتمال الحقيقي هو ١ / ٣ أكثر ممَّا إذا كان ١ / ٢ أو ١ / ١٠؛ ومن ثَمَّ يبدو من المعقول أن نختار القيمة ١ / ٣ كتقدير وحيد لاحتمال ظهور وجه الصورة؛ فهذه هي القيمة التي يرجح أن تسفر عن البيانات التي حصلنا عليها فعلًا.

يوضح هذا المثال طريقة «الإمكانية القصوى» للتقدير؛ إذ نختار قيمة المعلمة التي لديها أعلى احتمال لإنتاج البيانات المرصودة. في المثال السابق، حسبت فقط هذا الاحتمال للقِيَم الثلاث المرتبطة باحتمالية ظهور وجه الصورة (١ / ٣، ١ / ٢، ١ / ١٠)، ولكن يمكننا جوهريًّا حسابها لجميع القِيَم الممكنة. والدالة التي تبيِّن احتمال حدوث البيانات المرصودة لكل خيار ممكن لاحتمالية ظهور وجه الصورة يطلق عليها اسم «دالة الإمكان». وتلعب هذه الدالة دورًا محوريًّا في الاستدلال الإحصائي.

ويمكن تطبيق المبدأ نفسه من أجْل الحصول على تقديرات لمعلمات التوزيع الطبيعي، أو أي توزيع آخر. فنحسب ببساطة ما يمكن أن يكون احتمال الحصول على مجموعة بيانات مثل التي حدثت في الواقع بالنسبة لخيارات القيم المختلفة المحتملة للمعلمة. ومُقدر الإمكانية القصوى هو قيمة المعلمة التي تُنتج أكبر الاحتمالات. لاحِظ أن هذه العملية تُنتج قيمة واحدة؛ وهي تقدير يكون هو الأفضل من منظور الإمكانية القصوى. ولأنها قيمة واحدة فحسب، فإنها تسمى «تقدير النقطة».

ثمة طريقة بديلة للتفكير في هذا النهج للتقدير؛ وهي النظر لدالة الإمكان على أنها مقياس للتوافق بين البيانات المرصودة (التسلسل الناتج عن ست قذفات للعملة) والبيانات التي تنبَّأَت بها نظريتنا (حيث تعني كلمة «نظرية» هنا القيمة المقترحة لاحتمال ظهور وجه الصورة؛ على سبيل المثال، ١ / ٣ أو ١ / ٢). واختيار النظرية (احتمال ظهور وجه الصورة) لتحقيق أقصى قدر من التوافق — أو على نحو مكافئ، لتقليل التناقض — أمر معقول على نحو واضح. والتفكير في الأمر بهذه الطريقة يسمح لنا بالتعميم؛ إذ يمكننا التفكير في مقاييس أخرى للتناقض؛ على سبيل المثال، في كثير من الحالات، يتمثل مقياس جيد للتناقض في مجموع مربعات الفروق بين قيمة المعلمة

المقترحة وقيم العينة الفردية. واختيار المعلمة للحد من هذا المقياس يعني الحصول على «أفضل» تقدير، في سياق أصغر مجموع للفروق المربعة. في الواقع، هذه طريقة شائعة للغاية للتقدير، ويطلق عليها — لأسباب واضحة — «تقدير المربعات الصغرى».

أحيانًا يكون لدينا أفكار قبل تحليل البيانات عن القيمة التي نتوقع أن تكون عليها المعلمة. ومثل هذه الأفكار قد تأتي من الخبرات أو التجارب السابقة؛ على سبيل المثال، بناءً على خبرتنا السابقة في قذف القِطَع النقدية، ربما نعتقد أن المعلمة p، التي تعطي احتمال أن العملة المقذوفة سوف تظهر وجه الصورة، تقترب من ١ / ٢، وأنه من غير المحتمل جدًّا أن تكون بعيدة عن ١ / ٢. ونقول إن لدينا «توزيعًا قَبليًّا» لإيماننا بأن المعلمة المجهولة تأخذ قيمًا مختلفة. ويمثل هذا التوزيع إيمانًا ذاتيًّا حيال قيمة المعلمة؛ كما هي الحال مع التفسير الذاتي للاحتمال المذكور في الفصل الرابع. وفي مثل هذه الحالات، بدلًا من تحليل البيانات بمعزل لاستخراج تقدير لقيمة المعلمة، من المنطقي الجمع بين البيانات وإيماننا السابق لاستخراج «توزيع بعدي» لمعتقداتنا حول القيم المحتملة للمعلمة. وهذا يعني أننا نبدأ بتوزيع يصف معتقداتنا حول القيم المحتملة للمعلمة، ونعدِّله وفقًا لما نلاحظه في البيانات؛ على سبيل المثال، توزيعنا القبلي لاحتمال أن العملة ستظهر وجه الصورة ربما يكون مركَّزًا للغاية حول قيمة ١ / ٢؛ فنعتقد أنه من المحتمل جدًّا أن تقترب من ١ / ٢. ومع ذلك، إذا قُذفت العملة مائة مرة، وظهر في ثلاث مرات فحسب من أصل مائة مرة وجه الصورة، فربما نرغب في ضبط هذا التوزيع؛ بحيث تعتبر القيم الأصغر للاحتمال أكثر ترجيحًا والقيم الأقرب للقيمة ١ / ٢ أقل ترجيحًا.

في الواقع، نظرية بايز — المذكورة في الفصل الرابع — هي التي تمكننا من الجمع بين المعتقدات القَبلية والبيانات المرصودة لإنتاج المعتقدات البعدية. لهذا السبب، يطلق على هذه الطريقة للتقدير طريقة «التقدير البايزي». تذكر أن نظرية بايز تربط اثنين من الاحتمالات الشرطية: احتمال حدوث «أ» نظرًا لوقوع «ب»، واحتمال حدوث «ب» نظرًا لوقوع «أ». في هذه الحالة، نَستخدم النظرية لربط احتمال أن المعلمة لها قيمةٌ ما نظرًا للبيانات التي نلاحظها، مع احتمال ملاحظة هذه البيانات نظرًا لقيمة معينة للمعلمة. والآن، الاحتمال الثاني من هذين الاحتمالين — احتمال ملاحظة هذه البيانات نظرًا لقيمة معينة للمعلمة — هو ببساطة دالة الإمكان؛ ومن ثَمَّ تَستخدم نظرية بايز إمكانية البيانات لتعديل معتقداتنا القبلية، من أجل إنتاج معتقداتنا البعدية.

لاحظ أن هناك فرقًا دقيقًا — ولكنه مهم — بين هذه الطريقة والطرق الأخرى المذكورة سابقًا (التي غالبًا ما يُطلَق عليها الطرق «التكرارية» أو «الكلاسيكية»)؛ حيث إننا نفترض فيها أن المعلمة المجهولة لها قيمة ثابتة ولكنها مجهولة. ومع ذلك، بالنسبة للتقدير البايزي، افترضنا أن المعلمة المجهولة لها توزيع عبر مجموعة من القيم الممكنة، مقدم في البداية من خلال التوزيع القَبْلي، ثم بعد ذلك — عند تحديثه بواسطة المعلومات في البيانات — من خلال التوزيع البَعْدي. ويقر الباحث بأن المعلمة يمكن أن يكون لها قيم مختلفة، ويستخدم التوزيع الاحتمالي للتعبير عن معتقده حيال كل قيمة.

لا يخلو مفهوم التوزيع القبلي من الجوانب المُثِيرة للجَدَل. فعَلَى أقل تقدير، الأشخاص المختلفون ذوو الخبرة السابقة المختلفة، ربما يكون من المتوقَّع أن يمتلكوا توزيعات قبلية مختلفة، وهذه التوزيعات ستُجمع مع البيانات لإنتاج توزيعات بعدية مختلفة، وربما استنتاجات مختلفة. وهكذا تم التضحية بأي تظاهر بالموضوعية. كما توجد أيضًا صعوبة عملية؛ ففي حين أن متوسط التوزيع الطبيعي والمعلمة p في توزيع برنولي لهما تفسيرات واضحة ومباشرة، فليست الحال دائمًا أن تمتلك معلمات التوزيعات تفسيرات واضحة. ويمكن أن يكون أحيانًا من الصعب للغاية الوصول لتوزيعات قبلية معقولة تعكس معرفتنا المسبقة.

عند هذه النقطة في شرحنا لطريقة التقدير البايزي وصلنا إلى التوزيع البعدي؛ وهو توزيع يلخص اعتقاد الباحث بشأن كل قيمة تأخذها المعلمة بعد رؤية البيانات. ويمكننا، إذا أردنا، تقليص هذا التوزيع بأكمله لتقدير نقطة واحدة عن طريق استخدام ملخص إحصائي للتوزيع؛ على سبيل المثال، يمكننا أن نستخدم المتوسط أو المنوال الخاص به.

(٢) أي تقدير أفضل؟

كيف يمكننا معرفةُ ما إذا كانت طريقة تقدير النقطة فعالة أم لا، وأي مُقَدِّر هو الأفضل؟ على سبيل المثال، بينما قد أختار تقدير متوسط التوزيع باستخدام متوسط عينة مأخوذة من هذا التوزيع، ثمة بديل يتمثل في إسقاط أكبر القيم وأصغرها من العينة قبل احتساب المتوسط. وعمومًا، تتسم أكبر القيم وأصغرها بالقَدْر الأكبر من التفاوت من عينة لأخرى؛ لذلك ربما يُنتج التغاضي عنها تقديرًا أكثر موثوقية وأقلَّ تفاوتًا.

بالنسبة للطريقة التكرارية للتقدير، والتي تَفترض وجود قيمة حقيقية ثابتة — ولكنها مجهولة — للمعلمة الجاري تقديرها، نَوَدُّ في الحالة المثالية أن نعرف أيٌّ من هاتين الطريقتين تعطي تقديرًا أقرب إلى القيمة الحقيقية. وللأسف، بما أن القيمة الحقيقية مجهولة (بيت القصيد هنا هو تقديرها!) فلا يمكن أبدًا أن نعرف الإجابة. من ناحية أخرى، ما «يمكننا» أن نأمل في أن نعرفه هو عدد المرات التي قد نتوقَّع فيها أن تكون القيمة المقدَّرَة قريبة من القيمة الحقيقية إذا حدث أن كررنا عملية أَخْذ عينة من القياسات واحتساب التقدير. فرغم كل شيء، بما أن القيمة المقدرة تستند على عينة، فمِن المرجَّح أن القيمة المقدرة ستكون مختلفة إذا أخذت عينة مختلفة؛ وهذا يعني أن التقدير في حدِّ ذاته متغير عشوائي، يختلف من عينة لعينة أخرى. وبما أنه متغير عشوائي، فإن له توزيعًا. وإذا علمنا أن هذا التوزيع متجمع بإحكام حول القيمة الحقيقية، فربما نعتبر طريقةَ التقديرَ طريقةً جيدةً. بعبارة أخرى، إذا كنَّا نعرف أن طريقةً ما «عادة» ما تُسفِر عن تقدير يكون قريبًا للغاية من القيمة الحقيقية للمعلمة، فربما نعتبرها طريقة جيدة للتقدير. ومع أن هذا لا يُخبِرنا شيئًا عن حالتنا المحددة، فإنه سيكون لدينا ثقة في الطريقة على نحو مبرر. فعلى أي حال، إذا كنتَ على علم بأن شخصًا ما يتنبَّأ تنبؤًا صحيحًا في ٩٩٩ من كل ١٠٠٠ مرة، فإنك بالتأكيد ستَمِيل إلى الوثوق به في أي حالة معينة. أنت تفعل ذلك مع سائقي القطارات والطيارين والمطاعم، وما إلى ذلك؛ فأنت تعرف أن السائق والطيار نادرًا ما يقع في حادث، والمطعم نادرًا ما يقدم طعامًا مسمَّمًا، لذلك تكون سعيدًا بالمخاطرة بأنه «في هذه المرة» ستكون الأمور على ما يرام.

باستخدام هذا المبدأ، طُوِّرت عدة مقاييس مختلفة لتقييم طرق التقدير التكرارية البديلة. يتمثل أحد هذه المقاييس في «التحيز»، وهذا يُخبِرنا بمدى حجم الفارق بين القيمة الحقيقية للمعلمة والقيمة المتوسطة لتوزيع القيم المقدرة. وعلى وجه التحديد، إذا كان هذا الفارق يساوي صفرًا (أي إذا كان متوسط توزيع القيم المقدرة يساوي القيمة الحقيقية)، يقال إن المُقدِّر «غير متحيز».

على سبيل المثال، نسبة ظهور وجه الصورة نتيجة قذف العملة عدة مرات تكون مُقدِّرًا غير متحيز لاحتمال أن العملة ستستقر ووجه الصورة لأعلى؛ إذ إن القيمة المتوسطة لتوزيع هذه النسبة في التجارب المتكررة تساوي الاحتمال الصحيح لظهور وجه الصورة. وللتوضيح، افترض أن الاحتمال الحقيقي لاستقرار العملة ووجه الصورة لأعلى هو ٠٫٥٥؛ وهو أمر مجهول بالنسبة لنا، وأننا قذفنا العملة عشر مرات، وقدرنا

هذا الاحتمال عن طريق نسبة ظهور وجه الصورة. ربما تُسفِر القذفات العشر عن ست مرات لظهور وجه الصورة؛ وهي نسبة تبلغ ٠٫٦، أو ثلاث مرات؛ وهي نسبة تبلغ ٠٫٣، أو خمس مرات؛ وهي نسبة تبلغ ٠٫٥، وهكذا. وفي المتوسط (متوسط يُحسب من خلال تكرارات خيالية للقذفات العشر) ستكون النسبة ٠٫٥٥ لأن نسبة ظهور وجه الصورة هي مُقدِّر غير متحيز لاحتمال أن العملة سوف تستقر ووجه الصورة لأعلى.

وعمومًا، المقدر ذو التحيز الكبير لن يُنظر إليه على نحو مفضل مثل المقدر غير المتحيز. وفي المتوسط، من خلال تكرار التجربة، فإن المقدر ذا التحيز الكبير سوف يُسفِر عن قيمة مختلفة كثيرًا عن القيمة الحقيقية.

ثمة مقياس آخَر لتحديد جودة المقدر هو «متوسط مربع الخطأ»؛ فبالنسبة لأي قيمة مقدَّرة معينة يمكننا — إذا عرفنا قيمة المعلمة الحقيقية — حساب مربع الفارق (أي «مربع الخطأ») بين التقدير والقيمة الحقيقية. التربيع مُفيد لسبب واحد؛ وهو أنه يجعل كل الأرقام موجبة. وبما أن التقدير نفسه متغير عشوائي يختلف من عينة لعينة أخرى، فإن مربع الخطأ هو أيضًا كذلك. وبما أنه متغير عشوائي، فإن لديه توزيعًا، و«متوسط» مربع الخطأ ببساطة هو متوسط هذا التوزيع. ومتوسط مربع الخطأ الصغير يعني أنَّ — في المتوسط — مربعَ الفارق بين القيمة المقدَّرة والقيمة الحقيقية صغيرٌ. ولا يُنظر إلى المُقدِّر الذي يُعرَف أن لديه متوسطَ مربعِ خطأٍ كبيرًا بنظرة مفضَّلة مثل ذلك الذي لديه متوسطُ مربعِ خطأٍ صغيرٌ؛ إذ لن يثق المرء كثيرًا في أن قيمته قريبة من القيمة الحقيقية.

(٣) تقدير الفترة

عندما تناولنا بعض الملخصات الإحصائية الأساسية بالدراسة في الفصل الثاني، رأينا أنها تلخص على نحو جيد جدًّا عينة من القيم عن طريق متوسطها أو ملخص إحصائي وحيد آخر، ولكن هذا ترك الكثير مما هو مرغوب فيه. وتحديدًا، فشلت هذه الملخصات في إيضاح مدى انتشار قيم العينة حول هذا المتوسط. وعالجنا هذه المشكلة من خلال تقديم المزيد من الملخصات الإحصائية مثل المدى والانحراف المعياري، والتي أشارت إلى مدى تشتت قيم العينة.

ينطبق المبدأ نفسه على التقدير. حتى الآن تناولنا تقديرات النقطة، وهي تقديرات تتمثل في قيمة «واحدة» تمثل أفضل تقدير بمعنًى ما. وبديل ذلك هو تقديم مجموعة

من القيم — أي «فترة» — نثق في أنها تحتوي على القيمة الحقيقية. دعنا نَعُدْ إلى صفقة العشرة/الخمسة جنيهات التي عرضها صديقي. سعَيْنا سابقًا للوصول إلى أفضل تقدير وحيد لاحتمال أن قذفة العملة ستُظهِر وجه الصورة. بدلًا من ذلك، يمكن أن نسعى للوصول إلى مجموعة من القيم التي نَثِق في أنها تشمل الاحتمال الحقيقي. ربما يمكننا أن نكون واثقين للغاية في أن الاحتمال الحقيقي يكمن بين ١ / ٤ و٢ / ٥، مثلًا. وهذا مثال على «تقدير الفترة».

وبما أن القيمة الحقيقية مجهولة، فلا نستطيع أن نقول على وجه اليقين إذا كانت أي فترة معينة سوف تشتمل في الواقع على القيمة الحقيقية أم لا. ولكن تخيَّلْ تكرار التمرين مرارًا وتكرارًا باستخدام عينات عشوائية مختلفة (تمامًا كما تخيلنا عندما حددنا التحيز سابقًا). يمكننا حساب تقدير الفترة لكل عينة من هذه العينات، وإذا أُنشِئت الفترات على نحو صحيح، فمن الممكن أن نقول إن نسبة معينة من الفترات (على سبيل المثال ٩٥٪ أو ٩٩٪ أو ما نختار) تشمل القيمة الحقيقية المجهولة.

بالعودة إلى عملة صديقي، لا نستطيع أن نقول على وجه اليقين إن أي فترةٍ معينةٍ، محسوبةٍ لأي عينة بيانات معينة، ستحتوي على الاحتمال الصحيح بأن العملة سوف تظهر وجه الصورة. ولكن يمكننا القول إن ٩٥٪ (أو ما نختار) من الفترات ستحتوي على الاحتمال الحقيقي. وبما أن ٩٥٪ من الفترات سوف تحتوي على القيمة الحقيقية، فإننا يمكن أن نثق على نحو كبير أن الفترة الواحدة التي حسبناها، استنادًا إلى العينة التي حصلنا عليها فعلًا (ص – ك – ص – ك – ك – ك في المثال) ستشمل القيمة الحقيقية؛ ولهذا السبب، تسمى هذه الفترات «فترات الثقة».

بالتحول إلى طرق التقدير البايزي، رأينا أن نتيجة التحليل البايزي هي توزيع بَعديٌّ كامل للقيم، وهذا التوزيع يخبرنا بقوة اعتقادنا في أن المعلمة لديها أي قيمة معينة. يمكن أن نترك الأمور عند هذا الحد؛ فعلى سبيل المثال، إذا كان للتوزيع انحراف معياري صغير فإن هذا يعني أننا كنَّا على ثقة كبيرة بأن قيمة المعلمة تكمن في نطاق ضيق. لكن في بعض الأحيان، من المريح تلخيص الأمور بطريقة مماثلة لفترات الثقة أعلاه، وتقديم فترة محددة بأكبر وأصغر قيمة؛ على سبيل المثال، يمكننا إيجاد فترة تحتوي على ٩٥٪ من المساحة الموجودة تحت التوزيع الاحتمالي البعدي داخلها. وبما أن التوزيعات تمتلك درجة من تفسير المعتقد، فإن هذه الفترات يمكن تفسيرها على أنها تُعطِي احتمال أن القيمة الحقيقية تكمن في داخلها. ولتمييزها عن فترات الثقة التكرارية، تُسمَّى هذه الفترات «فترات المصداقية».

(٤) الاختبار

يَستخدم الإحصائيون عبارتَي: «اختبار الفرضية» و«اختبار الدلالة» لوصف عمليتَي استكشافِ ما إذا كانت المعلمات في النموذج تأخذ قيمًا محددة أو تقع في نطاقات معينة. وربما يعني ذلك في أبسط مستوياته اختبار معلمة واحدة فحسب؛ على سبيل المثال، يمكن أن نعرف أن ٥٠٪ من المرضى الذين يعانون من مرض معين يتعافَوْن بتناول العلاج المعياري، وقد نخمن أن تناول علاج جديد مقترح يشفي ٨٠٪ من هؤلاء المرضى. المعلمة الوحيدة التي نهتم باختبارها هي نسبة الشفاء بالنسبة للعلاج الجديد، وسنودُّ أن نعرف ما إذا كانت ٨٠٪ بدلًا من ٥٠٪.

من الحقيقي أن الناس مختلفون؛ فهم يختلفون من حيث العمر والجنس واللياقة البدنية وشدة المرض والوزن ومجموعة من الأشياء الأخرى؛ وهذا يعني أنه حتى عندما يتناول أشخاص متماثلون الجرعةَ نفسها من الدواء نفسه، فإن الاستجابة تختلف؛ فربما يُشفَى البعض ولا يُشفَى البعض الآخر. وفي الواقع، من الممكن للغاية أن تختلف استجابة المريض نفسه في الأوقات المختلفة وتحت الظروف المختلفة. النموذج المعقول لهذه الحالة ربما يتمثل في أن أي مريض يتناول دواءً لديه احتمال $p = 0.5$ للشفاء. وفي مثالنا، نعلم أن $p = 0.8$ في ظل العلاج المعياري، ونظن أن $p = 0.8$ في ظل العلاج الجديد.

في هذه المرحلة، لمعرفة النسبة التي تُشفَى عن طريق الدواء الجديد، فإن ما نودُّ القيام به هو إعطاء الدواء الجديد لمجموعة المرضى بأكملها الخاضعة للدراسة، تحت كل الظروف الممكنة، ونرى النسبة التي تُشفَى. هذا مستحيل على نحو واضح، وما يتعين علينا القيام به هو إعطاء الدواء لعينة من المرضى وحسب، ويمكننا بعد ذلك حساب نسبة الشفاء في العينة. للأسف، بما أننا نتعامل مع عينة فحسب، وليس جميع المرضى، فإن مجرد حقيقة شفاء ٨٠٪ — على سبيل المثال — من العينة أو ٦٠٪ أو ٩٠٪ أو أي نسبة، لا تعني بالضرورة أن هذه النسبة ستُشفَى في مجموعة المرضى بأكملها. فإذا أخذنا عينة مختلفة، فسنحصل على الأرجح على نتيجة مختلفة.

ومع ذلك فإن العينة المأخوذة من مجموعة يُشفَى فيها عمومًا ٥٠٪ فقط من المرضى عادةً ما تكون نسبة الشفاء فيها أقل من العينة المأخوذة من مجموعة تبلغ نسبة الشفاء فيها ٨٠٪ من المرضى.

وهذا يعني أننا يمكن استخدام حد — t مثلًا — بحيث لو لاحظنا أن نسبة الشفاء في العينة أقل من t سوف نرجح فرضية ٥٠٪، وإذا لاحظنا نسبة شفاء في العينة أكبر من t، فسوف نرجح فرضية ٨٠٪. وفي الحالة الثانية، نقول إن إحصائيات العينة تقع في «منطقة الرفض» أو «المنطقة الحرجة»؛ حيث إن نسبة الشفاء للعلاج المعياري — ٥٠٪ — قد «رفضت».

بالقيام بذلك، فإننا نخاطر بالوقوع في أحد نوعين من الأخطاء؛ فقد نقرر أن الدواء الجديد يشفي ٨٠٪ من المرضى في مجموعة المرضى الخاضعين للدراسة بأكملهم في حين أنه في الحقيقة يشفي ٥٠٪ فقط، أو قد نقرر أن الدواء الجديد يشفي ٥٠٪ من المرضى في مجموعة المرضى الخاضعين للدراسة بأكملهم في حين أنه في واقع الأمر يشفي ٨٠٪. ترتِّب طريقةٌ تسمى طريقة «نيمان-بيرسون» لاختبار الفرضية الأمورَ بحيث يكون احتمال الوقوع في كلا هذين النوعين من الأخطاء معروفًا، وصغيرًا بما فيه الكفاية ليعطينا ثقة في النتائج.

إليك كيفية عمل ذلك: نبدأ بوضع افتراض؛ إذ نفترض أن الدواء الجديد يشفي ٥٠٪ فقط من المرضى، ويسمى هذا الافتراض «فرضية العدم». تنص فرضية أخرى تُسمَّى «الفرضية البديلة» على أن الدواء الجديد يشفي ٨٠٪ من المرضى. باستخدام حسابات الاحتمال الأساسية نتمكن من معرفة نسبة العينات التي سوف تُظهر نسبة شفاء — عن طريق المصادفة — أكبر من أي t مختارة، إذا كان افتراض ٥٠٪ (فرضية العدم) حقيقيًّا. وعادة ما تُختار t بحيث إنه إذا كانت فرضية العدم حقيقية، فإن ٥٪ أو ١٪ فقط من المرات تتجاوز نسبة الشفاء في العينة t.

في هذه الحالة، عندما تكون فرضية العدم حقيقية (أي إذا كان ٥٠٪ فقط من المجموعة الخاضعة للدراسة بأكملها سيُشفى) وحصلنا في الواقع على نسبة شفاء في العينة أكبر من t — مما يؤدي بنا إلى اتخاذ قرار لصالح نسبة الشفاء الكلي البالغة ٨٠٪ — فربما نكون واقعين في النوع الأول من الأخطاء المذكورة آنفًا (وهو ما يسمى تقليديًّا «خطأ من النوع الأول»). وعادة ما يستخدم الرمز $\propto$ لتمثيل احتمال حدوث خطأ من النوع الأول. ويعني اختيارنا لقيمة t في المثال أن $\propto$ ثابتة لدينا عند ٠,٠٥ أو ٠,٠١ أو أي قيمة نختارها.

إذا لاحظنا نسبة شفاء في العينة أكبر من t، حينها إما أن تكون فرضية العدم حقيقية (النسبة الحقيقية البالغة ٥٠٪)، ويكون حدثٌ ذو احتمال ضعيف (معدل

العينة أعلى من t، يحدث باحتمال $\propto$) قد وقع، أو تكون فرضية العدم غير حقيقية. هذان هما الاحتمالان الوحيدان الممكنان، وهذا هو جوهر طريقة نيمان-بيرسون لاختبار الفرضية؛ فعن طريق اختيار t بحيث يكون $\propto$ صغيرًا بما فيه الكفاية (ويعتقد عمومًا أن ٠٫٠٥ و٠٫٠١ صغيران بما فيه الكفاية)، نشعر على نحو معقول بالثقة عند الإشارة إلى أن فرضية العدم ليست حقيقية؛ لأنه لو كانت حقيقية لوقع حدث غير مرجح.

أما النوع الثاني من الأخطاء (يسمى بطبيعة الحال «خطأ من النوع الثاني») فينشأ عندما تكون الفرضية البديلة حقيقية (نسبة ٨٠٪ في المثال)، ولكن نسبة الشفاء المرصودة في العينة أقل من t. بما أننا اخترنا t للسيطرة على احتمال الوقوع في الخطأ من النوع الأول، لا يمكننا أن نختار t أيضًا للسيطرة على احتمال الوقوع في الخطأ من النوع الثاني. ومع ذلك، يمكننا أن نجعل احتمال الوقوع في الخطأ من النوع الثاني صغيرًا كما نشاء عن طريق أخذ عينة كبيرة بما يكفي. وهذا مرة أخرى هو تأثير قانون الأعداد الكبيرة؛ فزيادة حجم العينة يقلل من نطاق التفاوت في تقدير العينة؛ ومن ثَمَّ يقلل من احتمال أن يكون تقدير العينة أقل من t عندما تكون القيمة الحقيقية للمجموعة الخاضعة للدراسة بأكملها أعلى؛ أي عند قيمة ٨٠٪. وبالتحديد، من خلال جعل العينة كبيرة بما يكفي يمكننا أن نقلِّل مِن احتمال حدوث الخطأ من النوع الثاني إلى أي قيمة نراها مناسِبة. عادة ما يُستخدَم الرمز β لتمثيل احتمال حدوث الخطأ من النوع الثاني، ويستخدم مصطلح «القوة» لتمثيل $1 - \beta$؛ وهو احتمال اختيار الفرضية البديلة عندما تكون حقيقية.

إن موقف اختبار الفرضيات المذكور هنا يشبه الموقف في المحكمة، حيث يُفترض في البداية أن المتهم بريء (فرضية العدم)، وهنا يكون من الممكن حدوث نوعين من الأخطاء: الحكم على شخص بريء بأنه مذنب (النوع الأول) أو الحكم على شخص مذنب بأنه بريء (النوع الثاني).

لاحظ أن الفرضيتين تدخلان في طريقة نيمان-بيرسون لاختبار الفرضية: فرضية العدم والفرضية البديلة. في «اختبار الدلالة»، تخضع فرضية العدم فقط للاختبار؛ فالهدف هو «رفض» فرضية العدم إذا كانت القيمة الإحصائية الخاضعة للاختبار (نسبة الشفاء في العينة في المثال السابق) مختلفةً بما فيه الكفاية عمَّا يمكن توقُّعه في ظل فرضية العدم، أو «الفشل في رفضها» إذا لم تكن القيمة متطرفة للغاية. فلا توجد أي فرضية بديلة مذكورة بوضوح. ويستخدم المصطلح «قيمة p» لوصف احتمال أن

نرصد قيمة إحصائية خاضعة للاختبار متطرفة مثل تلك المرصودة في الواقع، أو أكثر تطرفًا إذا كانت فرضية العدم حقيقية.

وُضعت فكرتا فرضية العدم واختبار الدلالة من أجل مجموعة كبيرة من المشاكل، فثمة اختبارات معينة طُوِّرت وسُمِّيت في كثير من الأحيان باسم أحد مطوِّريها الأصليين (مثل اختبار والد، واختبار مان ويتني)، أو سُمِّيت تيمُّنًا بتوزيع الإحصائية المعنية الخاضعة للاختبار (مثل اختبار t، واختبار مربع كاي).

وتُعَدُّ اختبارات الفرضيات البايزية — ظاهريًّا على الأقل — أكثر وضوحًا؛ فوَفْقَ مبرهنة بايز، لدينا احتمالات بَعدية بأن كل فرضية حقيقية؛ ومن ثم نستطيع استخدامها لاختيار إحدى الفرضيات. وفي الممارسة العملية، فإن الأمور في بعض الأحيان تكون أكثر تعقيدًا.

(٥) نظرية القرار

وَصفتُ على نحو غير رسمي «الاختبار» بأنه معرفةُ ما إذا كانت معلمات نموذج تتخذ قيمًا معينة أو تقع ضمن نطاقات معينة. وهذا وصف جيد لكثير مما يدور في أي سياق علمي؛ فالهدف هو اكتشاف كيف تسير الأمور. ولكن في سياقات أخرى، مثل التجارة أو الطب على سبيل المثال، فإن الهدف عادة ليس مجرد اكتشاف قيمة المعلمات، ولكن الهدف هو التصرف وفق ما نحصل عليه من معلومات. فنريد أن ننظر إلى المريض، ونُجرِي عددًا من الملاحظات والتجارب، ونتخذ أفضل مسار للعلاج، وذلك باستخدام البيانات الناتجة. ربما يعني مصطلح «أفضل» أشياءَ كثيرةً مختلفةً، ولكن على نحو نظري، فإننا سوف نرغب في تعظيم الفائدة أو الربح أو «المنفعة»، أو على نحو مكافئ، تقليل التكلفة أو الخسارة. إذا كنَّا نستطيع تحديد «دالة منفعة» مناسبة، محدِّدين ما سيكون المكسب إذا طُبِّق كل فعل بينما تأخذ الحقيقة غير المعروفة كل قيمة من قيمها الممكنة، يمكننا عندئذٍ مقارنة «قواعد اتخاذ القرارات» المختلفة؛ أي الطرق المختلفة للاختيار بين الأفعال؛ على سبيل المثال، ربما نختار قاعدة اتخاذ القرار التي تزيد من الحد الأدنى للمكاسب التي يمكن جلبها، مهما كانت الحقيقة غير المعروفة. بدلًا من ذلك، إذا كنَّا نعمل ضمن إطار بايزي؛ ومن ثم كان لدينا توزيع بعدي للاحتمالات عبر الحالة غير المعروفة للحقيقة، يمكننا حساب متوسط قيمة الربح لكل قاعدة اتخاذ قرار، واختيار القاعدة ذات أكبر قيمة للمتوسط.

إليك مثالًا على ذلك. ربما ترغب شركةٌ ما في معرفة أيُّ مسار للعمل — إرسال رسالة أم إجراء مكالمة هاتفية — هو الأكثر فعالية في تشجيع عملائها على شراء أحدث منتجاتها. سيكون من غير الواقعي أن نتصور أن الإجراء نفسه سيكون أكثر فعالية لجميع أنواع العملاء؛ فسيستجيب بعض العملاء على نحو أفضل للرسالة، وسيستجيب البعض أفضل للمكالمة الهاتفية، ولكننا لا نعرف الوسيلة الأفضل لكل عميل. ولكن ربما تمتلك الشركة بيانات حول كل عميل؛ وهي المعلومات التي قدَّمها العميل عندما اشترى منها لأول مرة؛ البيانات التي تصف مشترياته السابقة، وما شابَه ذلك. باستخدام هذه البيانات، يمكننا صياغة قواعد لاتخاذ القرار، والتي تُخبِرنا بأمور مثل «إذا كان العميل يبلغ من العمر أقل من ٢٥ عامًا، ولديه نمط سابق من المشتريات العادية فقُمْ بإجراء «مكالمة هاتفية»؛ وخلاف ذلك قُمْ بإرسال «الرسالة».» ويمكن صياغة العديد من قواعد اتخاذ القرار المحتملة تلك. وبالنسبة لكل إجراء — مكالمة هاتفية أو رسالة — فإننا نستطيع تقدير الربح، ربما حتى من الناحية النقدية، إذا قمنا بهذا الإجراء واتضح أن العميل من النوع الذي يستجيب (أو لا يستجيب) جيدًا لهذا الإجراء؛ ومن ثم يمكن أن نختار قاعدة اتخاذ القرار التي تجعل الحد الأدنى للربح أكبر. أو يمكننا حساب متوسط توزيع العملاء من كل نوع، لإنتاج متوسط ربح لكل قاعدة اتخاذ قرار، ثم اختيار القاعدة التي تؤدي إلى أكبر متوسط ربح.

(٦) إذن أين نحن الآن؟

كان الاستدلال الإحصائي على مر السنين موضع جدل كبير، وأحيانًا كان الجدل محتدمًا للغاية. وعلى الرغم من أن طرق الاستدلال المختلفة تؤدي بالفعل أحيانًا إلى استنتاجات مختلفة، فإن التجارب تبيِّن أن الاستخدام البالغ الدقة لهذه الأساليب عن طريق إحصائيين يفهمونها جيدًا يؤدي عمومًا إلى استنتاجات متشابهة. هذا كله جزء من فن تطبيق الإحصاء ويدل على أن إجراء التحليل الإحصائي ليس مجرد ممارسة آلية للرياضيات؛ فهو يتطلب فهمًا للبيانات وخلفياتها، وكذلك فهمًا سليمًا لنظرية الاستدلال الأساسية.

تضع المدارس المختلفة للاستدلال الإحصائي درجات متفاوتة من التركيز على عدد من المبادئ المختلفة. ومن أمثلة هذه المبادئ «مبدأ الإمكان» (إذا امتلك نموذجان من النماذج المختلفة دالة الاحتمال نفسها، فإنه ينبغي أن يؤديا إلى النتائج نفسها)، و«مبدأ

أخذ عينات متكررة» (ينبغي تقييم الإجراءات الإحصائية على أساس كيف ستتصرف «في المتوسط» إذا طُبقت على العديد من العينات المتكررة)، و«مبدأ الكفاية» (المعني بتلخيص البيانات بحيث يتم إبقاء معلومات كافية لتقدير أي معلمة). يبدو كل مبدأ من هذه المبادئ معقولًا تمامًا، ولكنها ربما تتعارض أحيانًا.

كانت الأساليب التكرارية الكلاسيكية لسنوات عديدة هي الطرق الأكثر استخدامًا في الاستدلال، ولكنِ اكتسبت الأساليب البايزية شعبية كبيرة في السنوات الأخيرة. كان هذا نتيجة مباشرة لتطوير أجهزة الكمبيوتر القوية وأساليب الحوسبة الذكية، فضلًا عن الترويج بحماس لمثل هذه الأساليب من قِبَل مؤيديها؛ فالعلوم تُمارَس في سياق اجتماعي، والجوانب الإنسانية المتعلقة بكيفية انتشار وتراجع هيمنة الأفكار المختلفة للاستدلال على مدى العقود القليلة الماضية تُعد قصة رائعة.

ثمة نقطة أخيرة؛ آمل أن أكون قد أوضحتُ في هذا الفصل أن هناك جوانب مختلفة للاستدلال. وتحديدًا، ربما نكون مهتمِّين بمحاولة العثور على إجابات لأنواع مختلفة من الأسئلة. وتشتمل هذه الأسئلة على أسئلة مثل: بِمَ تخبرني البيانات؟ وماذا ينبغي عليَّ أن أُومِنَ به؟ وماذا ينبغي أن أفعل؟ وما إلى ذلك. وتتلاءم طرق الاستدلال المختلفة مع الأنواع المختلفة من الأسئلة.

الفصل السادس

النماذج والأساليب الإحصائية

أفضل شيء في كون المرء إحصائيًّا هو أنه يستطيع اللعب في الفناء الخلفي للجميع.

جون دبليو توكي

(١) النماذج الإحصائية: وضع اللَّبِنات معًا

استخدمتُ التعبير «نموذج إحصائي» في أماكن مختلفة في هذا الكتاب حتى الآن دون تحديد ما أعنيه. النموذج الإحصائي هو تمثيل أو وصف بسيط لشيء أو نظام يخضع للدراسة. وربما ينطوي النموذج البسيط للغاية على جانب واحد فحسب من الطبيعة. وفي الواقع، رأينا أمثلة على ذلك في الفصل الرابع عندما تناولنا توزيعات المتغيرات المفردة. وعمومًا، يمكن بالفعل أن تكون النماذج الإحصائية مفصَّلة للغاية؛ إذ ربما تحتوي على آلاف المتغيرات المرتبطة بطرق معقدة للغاية. وعلى سبيل المثال، سوف يستخدم الاقتصاديون الذين يحاولون توجيه قرارات أي بنك مركزي مثل هذه النماذج الكبيرة.

ثمة منظور مهمٌّ فيما يخص النماذج يتمثل في التساؤل ما إذا كانت هذه النماذج تمثل الواقع الأساسي على نحو صحيح؛ أي ما إذا كانت «حقيقية» أم لا. في الواقع، هذا هو المنظور الذي اتخذناه سابقًا في هذا الكتاب عندما سألنا ما إذا كانت قيمة المعلمة المقترحة هي القيمة الحقيقية أم لا. ومع ذلك، يقر المنظور الأكثر تطورًا أنه لا يوجد نموذج — إحصائي أو غير ذلك — يمكن أن يأخذ في الاعتبار كل التأثيرات والعلاقات

الممكنة في العالم الحقيقي. وهذا المنظور هو الذي دفع الإحصائي البارز جورج بوكس للتأكيد على أن «جميع النماذج خاطئة، وإن كان بعضها مفيدًا.» إننا نبني نماذج لسبب؛ وهو مساعدتنا في الفهم والتنبؤ واتخاذ القرار، وما إلى ذلك. ورغم أننا ندرك أن نماذجنا تمثل تبسيطًا ضروريًّا للتعقيد الرهيب للعالم، فإننا إذا ما اخترناها جيدًا فسوف تمكننا من القيام بهذه الأمور. أما إذا اخترناها على نحو سيئ، فلن نفهم، وسوف تُخفِق توقُّعاتنا، وسوف تؤدي قراراتنا إلى أخطاء؛ إذن، هدفنا هو بناء نماذج جيدة بما فيه الكفاية لتحقيق غرضنا.

ويمكن تقسيم النماذج الإحصائية على نحو ملائم إلى نوعَيْن، يُسمَّيَان غالبًا «النماذج الآلية» و«النماذج التجريبية». يَستَنِد النموذج الآلي على بعض النظريات الأساسية الصلبة لكيفية ارتباط الأشياء؛ على سبيل المثال، ربما تُخبِرنا نظريةٌ ما في الفيزياء كيف أن سرعة سقوط الأجسام تزيد مع زيادة الزمن الذي تقع فيه. أو ربما تخبرنا نظرية أخرى حول كيفية انتشار العقاقير في أنحاء الجسم. في كلتا هاتين الحالتين، سوف تستند النماذج إلى نظريات حول كيفية عمل الأشياء فعليًّا؛ في الواقع، سوف تستند النماذج على المعادلات الرياضية التي تَصِفُ هذه النظريات، والبيانات التي نجمعها لتقييم نماذجنا سوف تكون قِيَم المتغيرات المستخدمة في هذه النظريات، مثل السرعة والزمن (في حالة سقوط الشيء) والتركيز والزمن (في حالة انتشار العقاقير)؛ ومن ثَمَّ النماذج الآلية هي طرق رياضية مباشِرة لوصف النظريات.

في المقابل، النماذج التجريبية هي مجرد محاولات لتوفير ملخصات ملائمة للجوانب المهمة من البيانات المرصودة. قد لا يكون لدينا أي نظرية تقول إن الأجسام الساقطة تزيد سرعتها مع مرور الزمن، ولكننا قد نلاحظ وجود علاقة بين الزمن والسرعة، وعلى أساس هذا، نُخمِّن وجود علاقة طردية. وإذا لم يوجد أي قاعدة نظرية أساسية لهذه العلاقة المقترحة، فإن النموذج يكون نموذجًا تجريبيًّا.

النماذج الآلية واسعة الانتشار في العلوم الفيزيائية وفي مجالات مثل الهندسة، فيما تميل العلوم الاجتماعية والسلوكية إلى الاستفادة على نحو أكبر من النماذج التجريبية. ومع ذلك فمن الواضح وجود تداخل كبير؛ إذ إن طبيعة النموذج تعتمد على ما يجري نَمْذجته ومدى سهولة فهمه؛ فالاقتصاد — الذي يُعَدُّ علمًا اجتماعيًّا — مليء بالنماذج الآلية المعتمِدة على نظريات حول كيفية ارتباط العوامل الاقتصادية. وعمومًا، ربما من الإنصاف القول إنه في المراحل الأولية لاستكشاف ظاهرةٍ ما، فإن النماذج التجريبية

تكون أكثر شيوعًا؛ إذ إن المرء يبحث عن الاتساق والأنماط في مجموعة الملاحظات. وفي مراحل لاحقة، عندما يكون الفهم قد ازداد، تُصبِح النماذج الآلية أكثر أهمية. وعلى أي حال، كما توضح نماذجنا للأجسام الساقطة، يمكن بناء نموذج معين على أنه نموذج تجريبي ثم يصبح آليًّا عندما يزداد فهمنا للظاهرة.

أحيانًا ما يكون من المفيد التمييز بين مختلف الاستخدامات الممكنة للنماذج الإحصائية. أحد أمثلة هذا التمييز يكون بين «الاستكشاف» و«التأكيد»؛ ففي الاستكشاف، نبحث عن العلاقات أو الأنماط؛ بينما في التأكيد، نهدف إلى معرفة ما إذا كانت البيانات تدعم تفسيرًا مقترحًا أم لا؛ لذلك، على سبيل المثال، في دراسة استكشافية ربما نبحث عن المتغيرات التي ترتبط معًا ارتباطًا وثيقًا. فربما يأخذ متغير واحد قيمة عالية كلما فعل ذلك متغير آخر، أو ربما تأخذ مجموعات من المتغيرات قيمًا متشابهة جدًّا مع أشياء مختلفة، وما إلى ذلك. من ناحية أخرى، ربما نستخدم البيانات في الدراسات التأكيدية لتقدير معلمات نموذج إحصائي مقترح وإجراء اختبار إحصائي لمعرفة ما إذا كان التقدير قريبًا بما فيه الكفاية ممَّا توقعتْه نظريتنا. أصبحتِ الأساليب الإحصائية لاستكشاف البيانات ذات أهمية متزايدة في السنوات الأخيرة، مع تراكم مجموعات من البيانات أكبر وأكبر. وينطبق هذا على التطبيقات العلمية (مثل فيزياء الجسيمات وعلم الفلك)، وكذلك التطبيقات التجارية (مثل قواعد البيانات التي تحتوي على تفاصيل المشتريات من المتاجر، أو المكالمات الهاتفية، أو بيانات تدفق النقر على الإنترنت).

ثمة تمييز آخر مهم في النمذجة الإحصائية بين «الوصف» و«التنبؤ»؛ فعند وصف مجموعة من البيانات، يتمثل الهدف في تلخيصها بطريقة مريحة؛ على سبيل المثال، إذا كانت مجموعة البيانات تتكون من ملاحظات لعشرة متغيرات (الطول والوزن والزمن المستغرق في التوجه للعمل، وما إلى ذلك) لكل شخص من مليون شخص، فسنحتاج لكي نبدأ في فهمها إلى تقليل حجمها إلى حجم معقول؛ على سبيل المثال، يمكننا تلخيصها من خلال المتوسط الحسابي والانحرافات المعيارية لكل متغير، وكذلك عن طريق قياسات مدى ترابطها. حينها سيكون لدينا بعض الأمل في فهْمِ ما يجري حيث إننا وصفنا الخصائص العامة للبيانات على نحو مريح. وبالإشارة إلى هذا، كما رأينا في الفصل الثاني، فإن هذه الملخصات الوصفية لا تخلو من المخاطر. فإنها، بحكم طبيعتها، تبسط التعقيد الهائل لمجموعة البيانات بأكملها؛ لذلك يجب أن ننتبه لاحتمال أنَّ وصفنا الموجز أغفل شيئًا مهمًّا؛ على سبيل المثال، ربما فشل نموذجنا في الوضع في الاعتبار حقيقةَ

وجود مجموعتَين وراثيتين متميزتين في المجموعة الكاملة الخاضعة للدراسة؛ لذلك يلزم وجود نموذج أكثر تفصيلًا لتمثيل ذلك.

أما هدفنا في التنبؤ فهو استخدام بعض المتغيرات للتنبؤ بقيم متغيرات أخرى؛ على سبيل المثال، قد يكون لدينا مجموعة من البيانات التي تبيِّن تفاصيل النظام الغذائي في الطفولة لعينة من الأشخاص وطولهم بعد البلوغ. يمكننا باستخدام هذه البيانات بناء نموذج يربط الطول بعد البلوغ بالنظام الغذائي في الطفولة، ثم نستخدم النموذج للتنبؤ بالطول المستقبلي المحتمل لطفل يتبع نظامًا غذائيًّا معينًا. لاحظ أن جانبًا أساسيًّا من البيانات لازمٌ لهذه النماذج؛ إذ إننا نحتاج لِقِيَمٍ لكلٍّ من المتغيرات المتنبِّئة والمتغير المتنبَّأ به من عينتنا. وسوف يتضح أن هذا تمييز مهم جدًّا بين النماذج التنبُّئية والنماذج الوصفية، كما سنرى فيما يلي:

ومرة أخرى، ليس التمييز واضحًا دائمًا وضوح الشمس، فربما نكون ببساطة مهتمِّين بوصف العلاقة بين النظام الغذائي في الطفولة والطول بعد البلوغ، مع عدم وجود نية لاستخدام النموذج للتنبؤ بأحدهما عن طريق الآخر.

يوجد نوع آخر مهم من التنبؤ هو «التوقع»، وفيه نستخدم بيانات من الماضي لبناء نموذج يمكن استخدامه كأساس للتنبؤ بالقيم المحتملة لملاحظات لم تُرصد بعدُ؛ على سبيل المثال، ربما نفحص النمط الشهري لمبيعات أجهزة التلفاز على مدى السنوات الخمس الماضية، ونقدِّر استقرائيًّا نزعةَ المبيعات والتفاوت الموسمي من أجْل توقُّع المَبِيعات المحتملة خلال الاثني عشر شهرًا التالية.

للنماذج الإحصائية استخدامات أخرى أيضًا. تعرفنا سريعًا على دورها في اتخاذ القرار في الفصل الخامس، كما رأينا أيضًا في الفصل عينه كيف قُدِّرت معلمات التوزيعات. يتم ذلك عن طريق تحديد مقياس للتناقض بين البيانات المرصودة والتوزيع النظري، ثم اختيار قيمة المعلمة المقدَّرة التي تقلل قياس التناقض لأدنى حد. ويستمد مقياس شائع للتناقض من الإمكان، والذي يقيس مدى احتمال أن بيانات مثل البيانات المرصودة ستنشأ إذا أخذتِ المعلمات قيمًا مختلفة متعددة. والآن، بما أن التوزيعات هي أشكال بسيطة فحسب من النموذج، فإن المبادئ نفسها بالضبط تنطبق عند تجربة نماذج أكثر تفصيلًا (مثل تلك المذكورة فيما يلي). ومع ذلك، تنشأ ظاهرة غريبة بينما تصبح النماذج أكثر تفصيلًا.

سأذكر مثالًا بسيطًا للتوضيح؛ لنفترض أننا نريد بناء نموذج للتنبؤ بالرواتب الأولى للخريجين، استنادًا إلى البيانات التي تَصِف دراستهم، والمواد التي درسوها في الجامعة،

ونتائج امتحاناتهم، وأيضًا عوامل مثل العمر والجنس ومكان الإقامة، وما إلى ذلك. افترض أننا جمعنا عينة مكونة من مائة من الخريجين الجدد وجمعنا البيانات منها. عمومًا، إذا حاولنا أن نبني توقعاتنا على عدد قليل جدًّا من المتغيرات (مثل العمر فقط) فإننا لن نحصل على تنبؤات دقيقة للغاية؛ فالعمر، في حد ذاته، وحده لا يحتوي على معلومات كافية للسماح لنا بأن نعرف كم سيكون راتب الشخص المتخرِّج في الجامعة بدقة متناهية. لتحسين دقة التنبؤ فإننا بحاجة إلى إضافة المزيد من العوامل المتنبِّئة (مثل استخدام العمر ومجال الدراسة ودرجات الامتحان للتنبؤ براتب الشخص المتخرج). ومع ذلك — وهنا تبرز المعضلة — إذا أضفنا عددًا أكبر مما يلزم من المتغيرات المتنبِّئة فإن دقة التنبؤ للمجموعة الكاملة الخاضعة للدراسة ستقل؛ فعلى الرغم من أننا نستخدم مزيدًا من المعلومات حول الخريجين، فإن نموذجنا ليس جيدًا.

يبدو هذا مناقضًا للمنطق؛ فكيف يمكن لإضافة «مزيد» من المعلومات أن تؤدي إلى تنبؤات «أسوأ»؟

الجواب مراوغ، ويُطلَق عليه أسماء مختلفة، منها الاسم المُعبِّر «الإفراط في المطابَقة». لفهم ذلك، دعنا نتراجع خطوة إلى الوراء ونتدبر هدفنا الحقيقي. إن هدفنا «ليس» الحصول على أفضل التنبؤات الممكنة للخريجين المائة في عيِّنتنا؛ فنحن نعلم بالفعل رواتبهم الأولى، ولكنَّ هدفنا هو الحصول على أفضل التنبؤات الممكنة بالنسبة للخريجين الآخرين؛ أي إن هدفنا هو «التعميم» من العينة الموجودة لدينا. والآن، بإضافة المزيد والمزيد من المتغيرات المتنبِّئة، فإننا بالتأكيد نضيف معلومات سوف تمكِّننا من التنبؤ برواتب الأشخاص الموجودين في عينتنا بالفعل على نحو أكثر دقة. ولكن العينة ليست سوى عينة؛ أي إنها لا تمثل رواتب المجموعة بأكملها على نحو كامل. وبعد فترة من الوقت، وبينما نواصل إضافة المزيد من المتغيرات المتنبئة، نبدأ في التنبؤ بجوانب من البيانات خاصة بالعينة وحدها؛ فهي ليست سمات تنطبق على المجموعة الكلية بأكملها.

تنطبق هذه الظاهرة على جميع النماذج الإحصائية؛ فالنماذج يمكن أن تكون مفرطة في التعقيد، بحيث تتطابق مع البيانات المرصودة جيدًا جدًّا بالفعل، ولكنها تفشل في التعميم على أشياء أخرى مستمَدَّة من التوزيع نفسه؛ وهذا يعني أنه لا بد من وضع استراتيجيات لاختيار نماذج بدرجة تعقيد مناسبة؛ فإذا كانت النماذج مفرطة التبسيط، فإننا نخاطر بفقدان قدرتها على التنبؤ، وإذا كانت مفرطة التعقيد، فإننا

نخاطر بالإفراط في المطابقة. يشكل هذا المفهوم أساس مبدأ «شفرة أوكام»، الذي ينص على أن «النماذج ينبغي ألَّا تكون أكثر تعقيدًا مما هو ضروري» (ينسب إلى الراهب الفرنسيسكاني ويليام الأوكامي من القرن الرابع عشر).

ولمشكلة الإفراط في المطابقة أهمية خاصة في مجال علم الإحصاء الحديث؛ فقبل ظهور أجهزة الكمبيوتر، وقبل أن يصبح مألوفًا مطابقةُ النماذج المعقدة مع أعداد كبيرة من المعلمات، كان خطر الوقوع في الإفراط في المطابقة أقل.

(٢) الأساليب الإحصائية: تطبيق الإحصاء

الهدف من هذا الجزء هو تحديد بعض الفئات المهمة من الطرق الإحصائية، وإظهار كيفية ارتباط بعضها ببعض، وتوضيح أنواع المشاكل التي يمكن استخدامها لحلها.

لنبدأ بالإشارة إلى أننا نهتم في كثير من الأحيان بالعلاقات بين أزواج المتغيرات. هل خطر الإصابة بالنوبات القلبية يزداد مع زيادة مؤشر كتلة الجسم؟ هل الاحترار العالمي ناتج عن النشاط البشري؟ هل إذا ارتفعت البطالة ينخفض التضخم؟ هل تحسين مزايا السلامة في السيارة يزيد مبيعاتها؟ وما إلى ذلك. إذا كان متغيران مرتبطين بحيث إن القِيَم الأكبر لأحدهما تَميل إلى الارتباط بالقِيَم الأكبر للآخر، يقال إن المتغيرين «مرتبطان إيجابيًّا». وإذا كانت القيم الأكبر لأحدهما تَميل إلى الارتباط بالقيم الأصغر للآخر، يقال إنهما «مرتبطان سلبيًّا». والطول والوزن لدى البشر مرتبطان إيجابيًّا؛ فالأشخاص الأطول يميل وزنهم إلى أن يكون أثقل. لاحظ أن العلاقة ليست علاقة دقيقة؛ إذ يوجد أشخاص طوال القامة أخفَّاء الوزن (الأشخاص النِّحَاف) وأشخاص قصار القامة ثقال الوزن. ولكن في المتوسط عمومًا، يرتبط طول القامة بالوزن الأثقل. يمكننا أيضًا أن نرى من هذا المثال أن محض الارتباط بين متغيرين لا يعني أن أحدهما يسبب الآخر؛ فإلزام شخصٍ ما باتباع نظام غذائي مكون من كعك بالكريمة لزيادة وزنه من غير المرجح أن يؤدي إلى زيادة طوله، ووضعه على مِخلعة لإطالة جسده من غير المرجح أن يزيد وزنه. في الواقع، كان الخلط بين الارتباط والسببية مصدرًا لكثير من سوء الفهم على مر السنين. من المرجح أن تُظهِر عينة عشوائية من الأطفال الذين تتراوح أعمارهم بين ٥ و١٦ سنة وجود ارتباط إيجابي واضح بين القدرة على القراءة والقدرة على القيام بعمليات حسابية. ولكن من غير المرجح أن تسبب إحداهما الأخرى، بل المرجح أن التقدم في العمر هو السبب الشائع لكلتيهما؛ فالأطفال الأكبر سنًّا أفضل في القراءة والحساب.

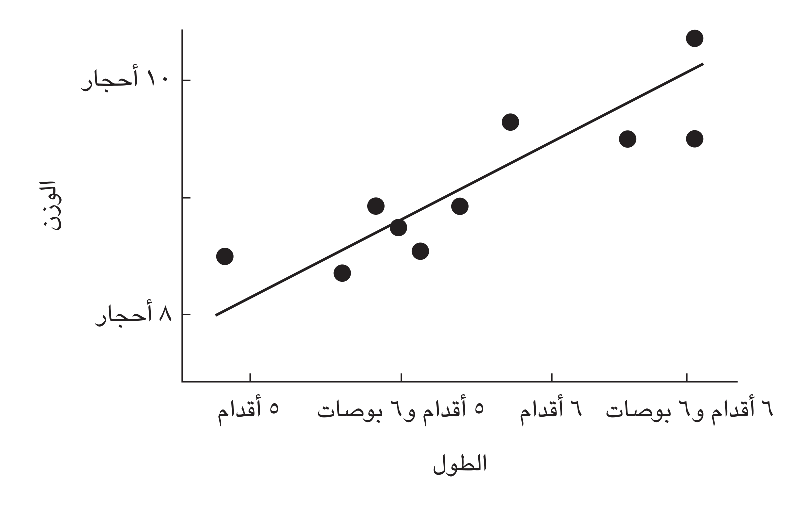
ثمة رقم واحد يمكن استخدامه لتمثيل قوة الارتباط، وهو «معامل الارتباط». ويوجد العديد من الطرق التي يمكن قياس هذه القوة بها، تمامًا مثلما رأينا أنه توجد طرق مختلفة لتعريف «المتوسط» و«التشتت». ومع ذلك، يوجد معيار عام لمعاملات الارتباط بأنها تَقَع بين −١ و+١؛ بحيث يعنى ٠ أنه لا يوجد ارتباط، ويعنى +١ وجود ارتباط إيجابي تام، ويعني −١ وجود ارتباط سلبي تام. ويعني الارتباط «التام» بين متغيرين «س» و«ص» أنك إذا كنتَ تعرف قيمة «س» فإنك تعرف قيمة «ص» بالضبط.

الارتباط علاقة متناظرة؛ فإذا كان الطول يرتبط بالوزن، فإن الوزن يرتبط بالطول، وقوة هذا الارتباط تظل نفسها مَهْمَا كانت الناحية التي ننظر إليها منها. وفي المقابل، نهتم في بعض الأحيان بالعلاقات غير المتناظرة بين المتغيرات؛ على سبيل المثال، ربما نرغب في معرفة مقدار الفرق في الوزن — في المتوسط — الذي يرتبط بوجود فارق في الطول يبلغ عشرة سنتيمترات. والإجابة على هذا النوع من الأسئلة تأتي من خلال طريقة إحصائية تسمى «تحليل الانحدار». ويخبرنا نموذج الانحدار بمتوسط قيمة المتغير «ص» لكل قيمة للمتغير «س». في المثال السابق، «انحدار الوزن على الطول» سيخبرنا بمتوسط الوزن الذي سيصل إليه الأشخاص عند كل طول. ويتضح هذا في الشكل ٦-١؛ حيث يمثَّل الوزن على المحور الرأسي، والطول على المحور الأفقي. وتوضح كل نقطة سوداء زوج الوزن/الطول لشخص من العينة. يبدو واضحًا الآن من هذا الشكل أننا لم نرصد قيمًا لجميع الأطوال الممكنة؛ على سبيل المثال، لا يوجد أي نقطة بيانات عند الطول الذي يبلغ بالضبط ٦ أقدام. إحدى طرق التغلب على هذه الصعوبة — بناء نموذج يعطينا متوسط وزن لكل قيمة من الطول — هي أن نفترض وجود علاقة بسيطة بين الطول ومتوسط الوزن. وهذه العلاقة البسيطة جدًّا هي علاقة خط مستقيم؛ ويَرِدُ مثال لهذا الخط في الشكل. وبالنسبة لأي طول معين، يسمح لنا هذا الخط بالبحث عن القيمة المقابلة من متوسط الوزن؛ فعلى سبيل المثال، وعلى وجه التحديد، فإنه يعطينا قيمة لمتوسط وزن الأشخاص الذين يبلغ طولهم ٦ أقدام.

وثمة عدة نقاط ينبغي توضيحها فيما يخص هذه الطريقة.

أولًا: إنها تعطي «متوسط» الأوزان عند كل طول. وهذا أمر معقول؛ إذ إنه في الحياة الواقعية، حتى الأشخاص ذوو الطول نفسه يمكن أن تتباين أوزانهم.

ثانيًا: نحن بحاجة إلى إيجاد طريقةٍ ما لتحديد الخط الذي نتحدث عنه بالضبط. يتضمن الشكل خطًّا واحدًا، ولكن كيف اخترنا هذا الخط وليس غيره؟ تتحدد الخطوط



شكل ٦-١: رسم خط وسط البيانات.

على نحو فريد عن طريق مَعلمتين — تقاطعهما (في هذا الشكل قيمة الوزن التي يتقاطع عندها الخط مع محور الوزن) وميلهما — لذلك نحن بحاجة إلى إيجاد وسيلة لاختيار هاتين المعلمتين أو تقديرهما. نعرف بالفعل طريقة تقدير المعلمة؛ فقد تناولناها في الفصل الخامس. ولتقدير المعلمات نختار تلك القيم التي تقلل من قدر التناقض بين النموذج والبيانات المرصودة. وبالنسبة لأي زوج معين (الوزن والطول) من البيانات، فإن أحد مقاييس التناقض هو مربع الفرق (مرة أخرى، السبب في كونه مربعًا هو جعل الأرقام موجبة) بين الوزن المرصود والوزن المتوقع عند هذا الطول. ويتمثل مقياس التناقض الكلي المعتمد على هذا في مجموع مربعات الفروق بين الأوزان المرصودة والأوزان المتوقَّعة عند الأطوال الواردة في البيانات. وبعد ذلك نقدر التقاطع والانحدار باختيار تلك القيم التي تقلل مجموع مربعات الفروق لأدنى درجة. وبما أنها تقلل (مجموع مربعات) الفروق بين القيم المرصودة والمتوقعة للأوزان في البيانات، فإن «خط انحدار المربعات الصغرى» هذا ينتِج أفضل تنبؤ لمتوسط الوزن عند أي قيمة للطول نختارها.

النقطة الثالثة: هي أنه على الرغم من أن هذا الافتراض بوجود علاقة خط مستقيم قد يبدو اعتباطيًّا إلى حدٍّ ما، فإنه مُبرَّر قليلًا. لماذا نختار خطًّا مستقيمًا، وليس خطًّا منحنيًا؟ دون الخوض في التفاصيل هنا، من الممكن تقديم منحنيات بدرجات متفاوتة بحيث يمكن أن يكون للخط الذي يبين العلاقة بين الطول ومتوسط الوزن أشكال أكثر تعقيدًا؛ فربما على سبيل المثال يزداد بسرعة أكبر عند الأطوال الأدنى من ازدياده عند الأطوال الأعلى. ونفعل ذلك من خلال جعل النموذج أكثر تعقيدًا، عن طريق إدخال معلمات إضافية بالإضافة إلى التقاطع والميل.

سعى مثال انحدار الطول/الوزن للتنبؤ بمتوسط الوزن من خلال متغير متنبئ واحد فقط هو الطول، لكن يمكننا أيضًا إدخال عوامل متنبئة محتملة أخرى من أجل تحقيق توقعات أكثر دقة؛ على سبيل المثال، يمتلك الرجال والنساء أشكال جسم مختلفة، بحيث إنه عند طول معين، ربما يكون الاختلاف في الأوزان بسبب نوع الجنس على نحو كبير؛ لذا يمكننا تضمين نوع الجنس أيضًا باعتباره عاملًا متنبئًا. ويمكننا مواصلة تضمين متغيرات أخرى نظن أنه من المرجح أن ترتبط بالوزن. لكن لا ينبغي أن نتمادى كثيرًا إذا كانتِ الملاحظات تتعلق بعدد محدد من الأشخاص فحسب، وإلَّا فسوف يتميز نموذجنا مرة أخرى بالإفراط في المطابقة مع البيانات؛ ولذا فإننا قد لا نرغب في تضمين كافة المتغيرات التي يمكن أن نفكر فيها، وإنما ندرج وحسب مجموعة فرعية منها.

بصفة عامة، ثمة أسباب أخرى أيضًا قد تدفعنا إلى الرغبة في تضمين مجموعة فرعية فقط من المتغيرات المتنبئة المحتملة؛ على سبيل المثال، ربما يكون قياس المتغيرات المتنبئة الإضافية مكلفًا، أو يستغرق وقتًا طويلًا؛ ولذا فإننا سوف نريد أن نُبقِي العدد عند أدنى حد ممكن. لهذه الأسباب وغيرها، طور الإحصائيون طرقًا للعثور على مجموعات فرعية جيدة من المتغيرات؛ حيث تعني كلمة «جيدة» أنها تنتج أفضل التنبؤات.

تربط نماذج الانحدار متغير ناتج أو متغير إجابة بواحد أو أكثر من المتغيرات المتنبِّئة. هذا نوع شائع جدًّا من المشكلات، وطُورت نماذج إحصائية أخرى للتعامل مع حالات مماثلة تختلف في بعض النواحي عن حالة الانحدار المستقيم؛ على سبيل المثال، في «تحليل البقاء» تُعرَف قيمة متغير الإجابة لبعض الحالات فقط، ويُعرف فقط أن قيمتها لحالات أخرى تتجاوز قيمةً ما. ينشأ هذا على نحو أكثر شيوعًا (على الرغم من أنه ليس في هذه الحالة وحسب) عندما يكون متغير الإجابة فترة زمنية؛ ومن ثَمَّ، فإننا قد نرغب في معرفة الفترة الزمنية التي سيظل فيها المريض على قيد الحياة (ومن هنا جاء

اسم هذه التقنية) أو طول الفترة الزمنية التي سيبقى فيها مكون من النظام قبل أن يحتاج إلى الاستبدال. وبأخذ الحالة الأولى كمثال للتوضيح، ربما تُبيِّن مجموعة البيانات المتوفِّرة لدينا أن أحد المرضى عاش خمسة أشهر، وعاش آخر شهرين فقط، وعاش ثلاثة آخرون أحد عشر شهرًا، وهكذا. ومع ذلك، ربما لم نتمكن لأسباب عملية من الانتظار حتى يموت آخِر مريض في الدراسة (الفترة التي قد تصل إلى أعوام)؛ لذلك توقفنا عن تسجيل الملاحظات. كل ما نعرفه عن بعض المرضى هو أنهم عاشوا فترة «أطول» من الوقت بين بدء رصد الملاحظات والتوقف عن رصدها. توصف هذه البيانات بأنها «مبتورة»، ولتوضيح التعقيدات التي تسببها، تأمَّلْ طريقة حساب متوسط فترة البقاء على قيد الحياة؛ فَلِحِساب المتوسط، نحتاج إلى جمع الفترات الزمنية المرصودة والقسمة على العدد الموجود. إننا لم نرصد في الواقع فترات البقاء على قيد الحياة للمرضى المبتورة بياناتهم، ولا يمكننا تضمينهم في الحساب. ولكن إذا أغفلناهم، فإننا سوف نُغفِل على وجه التحديد القِيَم الأكبر؛ لذلك سوف يكون تقديرنا متحيزًا إلى الأسفل. وعلى النقيض، إذا ضمَّنَّاهم، باستخدام فترات الملاحظة، فإن النتيجة تعتمد على وقت اختيارنا للتوقف عن رصد الملاحظات. وبما أن هذا غير ملائم أيضًا، فقد وُضعت أساليب أكثر تطورًا للتعامل مع البيانات المبتورة.

ثمة نسخة أخرى من مشكلة وجود متغير ناتج واحد مرتبط بواحد أو أكثر من المتغيرات المتنبئة تحدث في «تحليل التباين». يستخدم هذا التحليل على نطاق واسع في مجال الزراعة، وعلم النفس، ومراقبة الجودة الصناعية والتصنيع، وغيرها من المجالات. في تحليل التباين، تكون المتغيرات المتنبئة صريحة؛ وهذا يعني أن كلًّا منها يتخذ بضع قِيَم فحسب؛ على سبيل المثال، في تصنيع بعض المواد الكيميائية ربما نكون قادرين على السيطرة على درجة الحرارة والضغط والمدة، ويكون لدينا ثلاثة إعدادات لكلٍّ منها: منخفضة ومتوسطة وعالية. قابلنا مثل هذا الموقف عندما ناقشنا التصميم التجريبي في الفصل الثالث، وغالبًا ما يستخدم تحليل التباين لتحليل التجارب. ورغم تقديمه عادة على أنه مختلف عن تحليل الانحدار، فإنه من الممكن إعادة صياغته في صورة نموذج انحدار. وكلاهما حالتان خاصتان من فئة أكبر من النماذج تُسمَّى «النماذج الخطية».

وُسِّعت النماذج الخطية نفسها بطرق مختلفة. أحد التعميمات المهمة للغاية يتمثل فيما يسمى «النماذج الخطية المعمَّمة». في الانحدار وتحليل التباين، يكون الهدف هو التنبؤ بالقيمة المتوسطة للإجابة عند كل قيمة عامل متنبئ. وتوسِّع النماذجُ الخطيةُ

المعممةُ هذا من خلال السماح بكون غيرها من معلمات توزيع الإجابة، وليس المتوسط فقط، خاضعة للتنبؤ.

مع ذلك، تظهر نسخة أخرى من بنية الناتج/المتنبئ عندما تكون الإجابة نفسها قاطعة؛ على سبيل المثال، ربما تكون الإجابة عبارة عن قائمة من التشخيصات الطبية الممكنة، وربما تكون العوامل المتنبئة مزيجًا من الأعراض (قد تكون مدرجة على أنها حاضرة أو غائبة) ونتائج التحاليل الطبية. وتندرج هذه الأساليب تحت اسم عام هو «التصنيف المراقَب». وتحدث الحالة الخاصة الأهم من هذه النماذج عندما يكون متغير الإجابة ثنائيًّا؛ أي يأخذ قيمتين ممكنتين فحسب؛ مثل مريض/صحيح، مخاطرة جيدة/مخاطرة سيئة، مربح/عديم الجدوى، الكلمة المنطوقة «نعم»/الكلمة المنطوقة «لا» (في برامج التعرف على الكلام)، بصمة مصرح بها/بصمة غير مصرح بها (في أنظمة المقاييس الحيوية للتعرف على الأشخاص)، صفقة احتيالية/صفقة شرعية، وما شابه ذلك. وفي كل حال، فإن الهدف سيكون بناء نموذج يُمكِّننا من تحديد الفئة الأكثر احتمالًا للحالات الجديدة، مستخدمًا فحسب المعلومات في المتغيرات المتنبئة.

طور عدد كبير من الأدوات الإحصائية لمثل هذه الحالات. وكان من بين أول الأدوات «تحليل التمايز الخطي»، الذي طور في ثلاثينيات القرن العشرين، ولكنه لا يزال مستخدمًا على نطاق واسع للغاية حتى اليوم، سواء بشكله الأساسي أو بتوسيعاته الأكثر تفصيلًا. وتوجد طريقة أخرى تَحظَى بشعبية كبيرة في بعض المجالات — مثل الطب وإدارة قِيمة العُمَلاء — هي «تحليل التمايز اللوجستي». وهذا نسخة من الانحدار اللوجستي، وهو نوع من النماذج الخطية المعممة؛ لذلك يظهر الصلة الوثيقة بين طبقات الأدوات. في الواقع، يمكن اعتبار الانحدار اللوجستي أبسط أنواع «الشبكات العصبية». تُسمَّى الشبكات العصبية بهذا الاسم لأنها قُدِّمت في الأصل كنماذج لطريقة عمل المخ؛ إلَّا أنه في الوقت الحاضر تَركَّز العمل في هذا المجال كثيرًا على خصائصها الإحصائية كنظم للتنبؤ، بغض النظر عما إذا كانت تشكِّل نماذج جيدة للنظم الطبيعية أم لا.

وتوجد نماذج أخرى للتصنيف المراقب تشمل أسلوب «التصنيف الشجري» وطريقة «الجار الأقرب». يقسِّم النموذج الشجري المتغيرات إلى نطاقات، ويصنف نقاطًا جديدة وفقًا لمجموعة النطاقات التي تقع فيها. على سبيل المثال، ربما يُظهِر تحليل البيانات أن الأشخاص الذين تزيد أعمارهم عن ٥٠ عامًا ويعيشون نمط حياة قليل الحركة ولديهم مؤشِّر كتلة جسم أكبر من ٢٥؛ معرَّضون لخطر الإصابة بأمراض القلب. مثل

هذه النماذج يمكن أن تُمثَّل في صورة بنية شجرية؛ ومن هنا جاءت التسمية. في أسلوب الجار الأقرب، نجد الكائنات القليلة الموجودة في مجموعة البيانات التي تكون أكثر شبهًا (أو «أكثر قربًا») إلى الكائن الجديد الخاضع للتصنيف؛ حيث يتحدَّد التشابه من ناحية المتغيرات المتنبِّئة. بعدها يوضع الكائن الجديد ببساطة في الفئة نفسها كما هي حال غالبية هذه الكائنات المتشابهة كثيرًا.

ويسمى التصنيف المُراقَب بهذا الاسم لأنه يحتاج شخصًا (أي «مراقبًا») لتحديد تسميات فئات عينة البيانات، والتي يمكننا من خلالها بناء قاعدة التصنيف لتطبيقها على الكائنات الجديدة. ومع ذلك، لا يوجد في مسائل التصنيف الأخرى أي تسمية للفئات، والهدف هو ببساطة تقسيم الكائنات إلى فئات طبيعية، أو ربما فئات ملائمة. ويمكننا القول إن الهدف من ذلك هو تحديد الفئات؛ ففي الطب على سبيل المثال، ربما تكون لدينا عينة من المرضى لكلٍّ منهم تفاصيل عن أنماط الأعراض ونتائج التحاليل، وربما نظن أن عدة أنواع مختلفة من الأمراض ممثلة في العينة. سيكون هدفُنا حينها معرفةَ ما إذا كان المرضى يشكِّلون مجموعات مختلفة من منظور الأعراض ونتائج التحاليل. ويطلق على الأدوات الإحصائية لاستكشاف هذه التجمعات اسم «التحليل العنقودي». كان لهذه الأساليب فائدة كبيرة في تحديد الفرق بين الاكتئاب الأحادي القطب والثنائي القطب، وتستخدم في مجموعة كبيرة من المجالات الأخرى، منها — على سبيل المثال — إدارة قيمة العملاء والتسويق؛ حيث تكمن فائدتها في تحديدِ ما إذا كان يوجد أنواع مختلفة من العملاء أم لا.

في التحليل العنقودي، لا يوجد متغير «ناتج» ولا «إجابة». بدلًا من ذلك، فإن الهدف هو مجرد وصف البيانات على نحو سهل. وثمة أدوات إحصائية أخرى لها الهدف نفسه، على الرغم من أنها تسعى إلى وصف من نوع مختلف تمامًا؛ فعلى سبيل المثال، «النموذج البياني» هو وصف مبسط للعلاقات بين عدة متغيرات — وربما عدد كبير منها — استنادًا إلى افتراض أن العلاقات بين العديد من المتغيرات تسببها علاقات وسيطة مع متغيرات أخرى. وقد رأينا مثالًا بسيطًا جدًّا على هذا سابقًا؛ فربما كان الارتباط الإيجابي بين القدرة على القراءة والقدرة الحسابية لدى الأطفال نتيجة للعلاقة بين كلا هذين المتغيرين والعمر.

يمكن التوسع في هذه النماذج من خلال افتراض أن بعض العلاقات سببها المتغيرات «الكامنة» غير المقيسة التي تتعلق ببعض المتغيرات المرصودة؛ ومن ثَمَّ تحفز

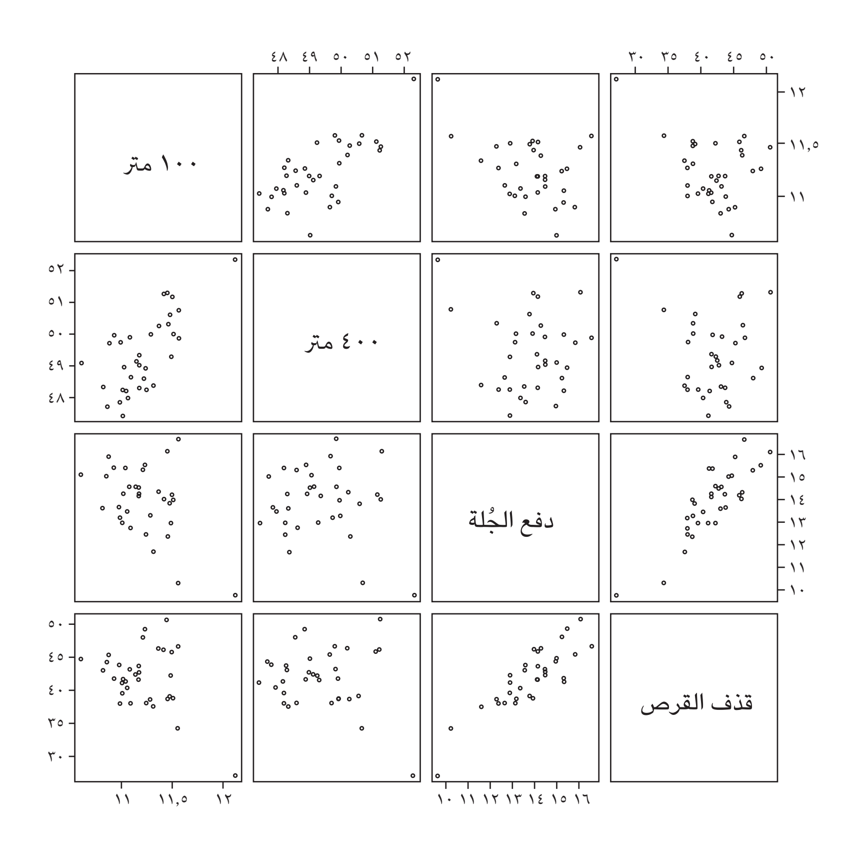
علاقة واضحة بينهما؛ فعلى سبيل المثال، ربما نلاحظ أن أسعار أسهم بعض الشركات ترتفع أو تنخفض معًا. إحدى طرق تفسير هذا قد تتمثل في تخمين وجود بعض المتغيرات الخفية (بعض جوانب الاقتصاد على سبيل المثال) التي ترتبط بكل سعر؛ ومن ثَمَّ تحفز العلاقة بين هذه الأسعار؛ فعندما يزيد المتغير الخفي، ترتفع كل الأسعار. تشكل هذه الأفكار أساس نماذج «التحليل العاملي»، وغالبًا ما يُسمَّى المتغير الكامن باسم «العامل الكامن». كما أنها تشكل أساس «نماذج ماركوف المخفية»، والتي فيها تُفسَّر سلسلة قيم مرصودة في سياق حالات خفية للنظام؛ على سبيل المثال، المرضى الذين يعانون من بعض الأمراض يتفاوتون من حيث جودة الحياة، فأحيانًا ينتكسون وأحيانًا يُشفَوْن على نحو مؤقت. ويمكن نمذجة هذا التعاقب في سياق الحالات الأساسية المتغيرة.

إذا كانت أساليب التصنيف سُمِّيت تيمُّنًا بأنواع المسائل المصمَّمة لحلها، فقد سميت أساليب أخرى تيمنًا بطبيعة البيانات التي تعمل عليها؛ على سبيل المثال، أساليب «تحليل السلاسل الزمنية» تعمل على السلاسل الزمنية؛ أي الملاحظات المتكررة للمتغير أو المتغيرات نفسها على مدار تسلسل زمني. وهياكل البيانات تلك موجودة في كل مكان؛ فهي توجد في الاقتصاد (مثل قياسات التضخم والناتج المحلي الإجمالي والبطالة)، والهندسة، والطب (مثل وحدات العناية المركزة)، وفي كثير من المجالات الأخرى. وفي تحليل السلاسل الزمنية، ربما يكون هدفنا هو فهمها، أو تحليلها إلى مكوناتها الرئيسية (مثل النزعة والموسمية)، أو رصد متى يتغير سلوك النظام، أو رصد الحالات الشاذة (مثل التنبؤ بالزلازل)، أو توقع القيم المستقبلية المحتملة، أو من أجل مجموعة من الأسباب الأخرى. وقد طورت مجموعة كبيرة من الأساليب لتحليل هذه البيانات.

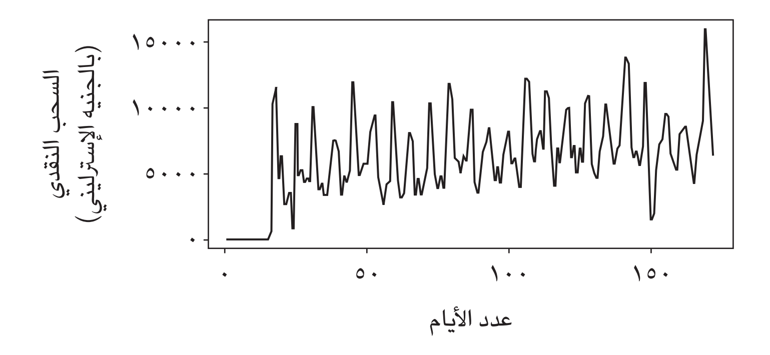
(٣) الرسوم البيانية الإحصائية

توجد فئة معينة من الأدوات الإحصائية مهمة للغاية لدرجة أنها تستحق اهتمامًا خاصًّا. وهذه الفئة هي استخدام الرسوم البيانية. صُقلت العين البشرية على مدار دهور من التطور لكي تكون قادرة على إدراك البِنَى والأنماط في الإشارات التي تَصِل إليها. ويستفيد علم الإحصاء استفادة مكثفة من ذلك عن طريق تمثيل البيانات في صورة مجموعة كبيرة من الأنواع المختلفة من الأشكال الرسومية؛ فعندما تُعرض البيانات على نحو جيد، فإن العلاقات بين المتغيرات أو التكوينات في البيانات تصبح واضحة. ويُستخدَم هذا في تحليل

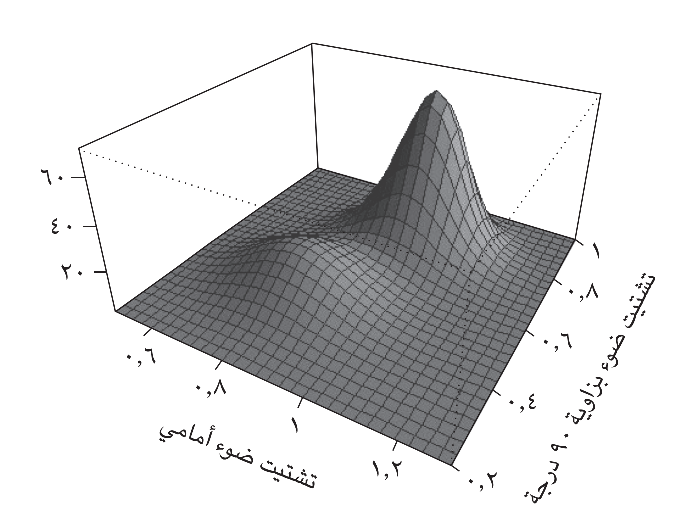
البيانات للمساعدة في فهمِ ما يدور (تَذكَّرْ توزيع رواتب البيسبول في الشكل ٢-١)، وإيصال النتائج إلى الآخرين. وأُقدِّمُ بعض الأمثلة في الأشكال الثلاثة التالية:



شكل ٦-٢: «مصفوفة الشكل الانتشاري» التي تُبيِّن أوقات سباق العَدْو ١٠٠ متر و٤٠٠ متر (بالثواني)، والمسافات (بالأمتار) لدفع الجُلة وقذف القرص للمنافسين في عشاري الرجال في دورة الألعاب الأولمبية عام ١٩٨٨. ويبيِّن كل مربع العلاقةَ بين اثنين من المتغيرات الأربعة. والترابط القوي بين نتائج حدثَيِ الرمي واضحٌ على نحو مباشر.



شكل ٦-٣: مخطط السلسلة الزمنية الذي يبين المبلغ المسحوب من جهاز صراف آلي كل يوم. يبين الشكل بوضوح وجود دورات أسبوعية وشهرية، وأيضًا وجود نزعة متزايدة تدريجيًّا مع مرور الوقت. ويتضح أيضًا وجود قيمة منخفضة على نحو مفاجئ بالقرب من نهاية الفترة.



شكل ٦-٤: توزيع قِيَم تشتيت الضوء من خلايا عوالق نباتية من أنواع مختلفة. في الواقع، يُعرض ثلاثة أنواع هنا، ولكن يمتلك اثنان منها توزيعين للقيم متشابهين جدًّا؛ لذلك يتجمع هذان التوزيعان لتكوين قمة عالية واحدة.

خاتمة

قَدَّم هذا الفصل مراجعة سريعة لعدد قليل من الأدوات الإحصائية المهمة، ولكن يوجد العديد من الأدوات الأخرى الرائعة التي لم أذكرها. وتتناسب النماذج المختلفة مع أنواع المسائل المختلفة وأنواع البيانات المختلفة، ويوجد عدد لا نهائي من المسائل وبنى البيانات. ومن المهم أيضًا أن ندرك أن النماذج ليست كيانات معزولة؛ فالحقيقة هي أن النماذج المختلفة ترتبط بطرق متعددة؛ فربما تكون النماذج تعميمًا لأنواع أخرى من النماذج أو تكون حالات خاصة منها أو تتكيف مع أنواع مختلفة من البيانات، بَيْدَ أنها مُدمَجَة جميعًا في شبكة غنية من العلاقات.

الفصل السابع

الحوسبة الإحصائية

السحر الحقيقي يأتي من فريق التحليل الإحصائي لدينا.

سام الخَلَف

(١) الإحصاء يغير تركيزه

رأينا في المناقشات السابقة كيف أن الإفراط في المطابقة يمكن أن يمثل مشكلة، لكننا لم نتطرق أيضًا إلى الحل؛ إذ إننا ببساطة أشرنا إلى أنه كان من الضروري اختيار نماذج ليست معقدة للغاية ولا بسيطة للغاية. وبدون امتلاك خبرة كبيرة في مجال النمذجة الإحصائية، ليست هذه نصيحة مفيدة جدًّا، وتوجد حاجة إلى مزيد من الطرق الموضوعية. وتستند إحدى هذه الطرق إلى مبدأ «التحقُّق المتبادل».

كما رأينا أنه — بصفة عامة — بينما يزداد تعقيد النموذج، تواصِل جودةُ مطابقته مع البيانات المتاحة التحسنَ، إلَّا أن جودة مطابقته مع عينات أخرى مستمَدَّة من التوزيع نفسه (أو «أدائه خارج العينة») تتحسن عادة في البداية، ولكنْ بعد ذلك تبدأ في التدهور. هنا تكون «العينات الأخرى» تمثيلًا للبيانات الجديدة، وهي ما نحن مهتمون به حقًّا. والنقطة التي يكون فيها النموذج مطابقًا على نحو أفضل مع بيانات «عينة أخرى» يبدو أن من شأنها أن تمنحنا نموذجًا ذا مستوًى مناسب من التعقيد. وهذا هو مفتاح الحل؛ فيجب علينا تقدير معلمات النموذج باستخدام عينة واحدة، وتقييم أدائه باستخدام عينة أخرى.

للأسف، عادةً ما نمتلك عينة واحدة فقط. وإحدى طرق مواجهة ذلك تتمثل في تقسيم هذه العينة (عشوائيًّا) إلى عينتين فرعيتين. وتستخدم عينة فرعية واحدة (تُسمَّى «عينة التدريب» أو «عينة التصميم») لتقدير المعلمة، وتستخدم الأخرى (تُسمَّى «عينة التحقق») لتقييم الأداء واختيار النموذج. وهذا هو أسلوب التحقق المتبادل. وفي العادة، لتخفيف أي مشاكل ناجمة عن كون العينة الفرعية المستخدمة لتقدير المعلمات ليست هي مجمل العينة الأصلية، يُكرر هذا الإجراء عدة مرات؛ يعني هذا أن العينة الأصلية تُقسَّم عشوائيًّا إلى عينتين، وتُقدَّر المعلمات باستخدام عينة فرعية واحدة، ويُقيَّم النموذج باستخدام الأخرى. ويُكرَّر هذا بتقسيمات عشوائية مختلفة للعينة. وأخيرًا، يُحسب متوسط نتائج تقييم كل التقسيمات، لكي يَنتُج قياسٌ عامٌّ للأداء المستقبلي المرجح.

يُعَدُّ التحقق المتبادَل مثالًا على نهج «مكثَّف حاسوبيًّا»؛ وسُمِّي هكذا للسبب الواضح المتمثل في ضرورة بناء نماذج متعددة. وتوجد فئة أخرى مهمة من هذه الأساليب هي «تقنية إعادة المعاينة»، ولهذه الطريقة مجموعة متنوعة من الاستخدامات، ولكنَّ أحد استخداماتها المهمة يتمثل في تقدير عدم اليقين المرتبط بالنماذج المعقدة؛ أي تحديد مدى الاختلاف الذي يمكننا أن نتوقع أن يصبح عليه النموذج إذا كنَّا قد أخذنا عينة بيانات مختلفة. وتعمل طرق إعادة المعاينة من خلال أخذ عينات فرعية عشوائية بحجم العينة الأصلية نفسها من العينة الأصلية (وهو ما يعني أن بعض نقاط البيانات ستستخدم أكثر من مرة). ويبنى نموذج جديد، بالشكل نفسه للنموذج الذي يجري تقييمه، لكل عينة من هذه العينات الفرعية. يبدو الأمر كما لو كان لدينا عينات متعددة، وكلها بالحجم نفسه، من التوزيع الأصلي، وتُنتج كلٌّ منها نموذجًا مُقدَّرًا. ويمكن بعد ذلك استخدام مجموعة النماذج تلك لمعرفة كيف كان يمكن أن يختلف هذا النموذج إذا كنَّا قد أخذنا عينة مختلفة.

أحد أقوى الأمثلة التوضيحية للكيفية التي غيَّرت بها قوة الكمبيوتر علم الإحصاء الحديث، يَظهر في تأثير الأساليب الكثيفة حاسوبيًّا على طرق الاستدلال البايزية المذكورة في الفصل الخامس. فمن أجْل استخدام الطرق البايزية عمليًّا، من الضروري حساب دوالِّ التوزيع المعقدة (بمصطلحات رياضية، توجد حاجة إلى تكامل عالي الأبعاد). وقد ساعدت أجهزة الكمبيوتر على تجنب هذه المشكلة؛ فبدلًا من تقييم التوزيعات رياضيًّا، يأخذ جهاز الكمبيوتر أعدادًا كبيرة من العينات العشوائية منها. ويمكن تقدير خصائص التوزيعات من هذه العينات العشوائية، بالطريقة نفسها لاستخدامنا لمتوسط العينة

لتقدير متوسط المجموعة الخاضعة للدراسة بأكملها. وأحدثت طريقة «مونت كارلو المستندة إلى سلسلة ماركوف» ثورة في ممارسة الإحصاء البايزية؛ إذ حوَّلتْها جوهريًّا من مجموعة من الأفكار الجذَّابة من الناحية النظرية، ولكنها قاصرة على النحو العملي إلى تقنية قوية لتحليل البيانات.

لَفَتَ الفصل السابق الانتباهَ إلى قوة الأساليب الرسومية البيانية، من أجل التوضيح وتوصيل الفكرة، ولكنْ نَقَلَ الكمبيوتر الأساليبَ الرسومية البيانية إلى مستوًى جديد تمامًا؛ فبينما لم يكن لدينا في الماضي سوى صور ثابتة بالأبيض والأسود، أصبح لدينا الآن صورًا ملوَّنة متحركة، بل وأهم من ذلك أننا يمكننا الآن التفاعل مباشرة مع الصورة. وكمثال بسيط فحسب، من الممكن عرض أشكال متعددة في الوقت ذاته، يبين كل واحد منها العلاقات بين أزواج مختلفة من المتغيرات المرتبطة بالكائنات، مثل مصفوفة الشكل الانتشاري في الشكل ٦-٢، ولكن في هذه الحالة ترتبط الأشكال من خلال الكمبيوتر. في هذه الحالة، إن إبراز أو تغيير أي مجموعة من النقاط يَظهر في الوقت نفسه في جميع الأشكال. وتسمح أدوات أخرى للمرء «بالطيران» على نحو تفاعلي خلال فضاء بيانات عالي الأبعاد، عارضًا البيانات بطرق متعددة.

وبما أن الإحصاء يُستخدَم على مستوًى عالمي، ولأن الكمبيوتر يلعب مثل هذا الدور المحوري، فإنه ليس من المستغرَب أن تُطوَّر حزم برامج إحصائية سهلة الاستعمال. ويُعَدُّ بعض منها مهمًّا لدرجة أنها أصبحتْ معايير في مجالات تطبيق معينة. ولكن هذا لا ينبغي أن يُنسِينا أن التطبيق الفعَّال للأدوات الإحصائية يتطلب تفكيرًا متأنِّيًا؛ ففي الواقع، في الأيام الأولى لتطوير البرمجيات الإحصائية، خَشِيَ البعض من أنَّ توافُر مثل هذه الأدوات من شأنه أن يزيل الحاجة للإحصائيين؛ حيث إنه «يمكن لأي شخص أن يقوم بالتحليل الإحصائي؛ فكل ما عليه القيام به هو إعطاء التعليمات المناسبة للكمبيوتر.» مع ذلك، ثبت أن العكس تمامًا هو الصحيح؛ وهناك مزيد من الطلب على الإحصائيين بمرور الوقت. وتوجد عدة أسباب لذلك.

أحد الأسباب هو أن البيانات تُسجَّل تلقائيًّا على نحو متزايد؛ ففي الحياة اليومية، في كل مرة تقوم فيها بإجراء عملية شراء ببطاقة الائتمان أو تتسوق في متجر، تُخزَّن تفاصيل العملية تلقائيًّا؛ وفي العلوم الطبيعية، تسجِّل الأدوات الرقمية الخواص الفيزيائية والكيميائية دون الحاجة إلى تدخل بشري؛ وفي المستشفيات، تراقب الأجهزة الإلكترونية المرضى تلقائيًّا؛ وما إلى ذلك. إننا نواجه سيلًا من البيانات. وهذا يمثل فرصة هائلة، ولكنْ يلزم وجود مهارات إحصائية للاستفادة منها.

السبب الثاني هو ظهور نطاقات جديدة تتطلب مهارات إحصائية؛ فالمعلوماتية الحيوية وعلم الجينوم يفككان التعقيد المُذهِل للجسم البشري من خلال البيانات التجريبية والرصدية، ويقومان على الاستدلال الإحصائي. وقد وُصف قطاع صناديق التحوط بأنه «قطاع مبني على الإحصاء»، وهو يستخدم الأدوات الإحصائية لوضع نماذج لسلوك الأسهم ومؤشرات الأسعار الأخرى.

السبب الثالث هو أن إعطاء الأوامر لجهاز كمبيوتر شيء، ومعرفة الأوامر التي ينبغي إعطاؤها وفهم النتائج شيء آخر تمامًا؛ فمِن المؤكَّد أن الأمر ليس مجرد مسألة اختيار الأداة المناسبة للوظيفة وترك الكمبيوتر يقوم ببقية العمل، بل الأمر يتطلب خبرة إحصائية وفهمًا. وبالنسبة للهُواة، من المهم أن يعرف المرء حدوده، ومتى يجب عليه طلب النصيحة من خبير إحصائي. وللأسف، تعرض وسائل الأعلام كل أسبوع أناسًا يتطرَّقون لأمور أكبر من فهمهم الإحصائي.

ولهذه الأسباب وأكثر، يشهد علم الإحصاء عصرًا ذهبيًّا.

وصلنا الآن إلى نهاية هذا الكتاب الموجَز. لقد رأَيْنا قدْرًا من التوسع غير العادي الذي يتسم به الإحصاء؛ إذ إنه يُطبَّق في معظم مناحي الحياة. ورأينا شيئًا من طُرُقه؛ الأدوات والمقاييس المتطورة التي يستخدمها. كما رأينا أيضًا أنه مجال ديناميكي، لا يَزَال ينمو ويتطور. ومع ذلك، قبل كل شيء، أرجو أن أكون قد أوضحتُ أن علم الإحصاء الحديث، المستنِد إلى الأسس الفلسفية العميقة، هو فن الاكتشاف؛ فعلم الإحصاء الحديث يمكننا من استخلاص أسرار الكون من حولنا؛ أي إنه يمكننا من الفهم.

تعليقات ختامية

إجابات لعبارات سوء الفهم الواردة في الفصل الأول:

(١) من الواضح أنه كلما كان اكتشاف المرض في وقت مبكر، طالتِ المدة التي سيعيشها المريض، بغضِّ النظر عن أي تدخل طبي؛ فبطريقة أو بأخرى يحتاج هذا إلى أن يؤخَذ بعين الاعتبار.

(٢) يعني التخفيض بنسبة ٢٥٪ أن السعر خُفِّضَ بمقدار الربع، ولكن هذا يعني أنه للعودة إلى السعر الأصلي عليك زيادة السعر بمقدار الثلث (٣٣٪)، وليس الربع (٢٥٪)؛ على سبيل المثال، الخصم البالغ ٢٥٪ على السعر الأصلي ١٠٠ جنيه استرليني يؤدي إلى السعر المُعلَن ٧٥ جنيهًا استرلينيًّا. وللعودة إلى السعر الأصلي علينا زيادة هذا السعر بمبلغ ٢٥ جنيهًا استرلينيًّا؛ أي ٣٣٪ من ٧٥ جنيهًا استرلينيًّا.

(٣) هذا يفترض أن متوسط العمر المتوقَّع سوف يستمر في الزيادة بالمعدل نفسه لزيادته في الماضي.

(٤) إذا كان طفل واحد قد قُتِل في عام ١٩٥٠، فإن العبارة تعني أن اثنين لَقِيَا مصرعهما في عام ١٩٥١، وأربعة في عام ١٩٥٢، وثمانية في عام ١٩٥٣، وستة عشر في عام ١٩٥٤، وما إلى ذلك. واستمرار المضاعفة بهذه الطريقة يعني أنه بحلول الوقت الراهن يُقتَل من الأطفال رميًا بالرصاص سنويًّا عدد أكثر من عدد سكان العالم. (وهذا المثال مأخوذ من الكتاب الممتاز الذي ألَّفه جويل بيست، والوارد في قسم القراءات الإضافية.)

قراءات إضافية

الفصل الأول

A. R. Jadad and M. W. Enkin, *Randomised Controlled Trials: Questions, Answers and Musings,* 2nd edn. (Malden, Massachusetts: Blackwell Publishing, 2007).

Joel Best, *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists* (Berkeley: University of California Press, 2001).

John Chambers, Greater or lesser statistics: a choice for future research, *Statistics and Computing,* 3 (1993): 18–24.

Foundation for the Study of Infant Death. (http://www.fsid.org.uk/cot-death.html). Accessed 6 April 2007.

Helen Joyce, Beyond reasonable doubt, *Plus Magazine* (2002). (http://www.plus.maths.org/issue21/features/clark/index.html). Accessed 14 July 2008.

(http://www.sallyclark.org.uk/). Accessed 14 July 2008.

الفصل الثاني

D. J. Hand, *Information Generation: How Data Rule Our World* (Oxford: Oneworld, 2007).

F. Daly, D. J. Hand, M. C. Jones, A. D. Lunn, and K. McConway, *Elements of Statistics* (Harlow: Addison-Wesley, 1995).

الفصل الثالث

S. Benvenga, Errors based on units of measure, *The Lancet,* 363 (2004): 1368.

T. L. Fine, *Theories of Probability: An Examination of Foundations* (New York: Academic Press, 1973).

الفصل الرابع

D. R. Cox, *Principles of Statstical Inference* (Cambridge: Cambridge University Press, 2006).

H. S. Migon and D. Gamerman, *Statistical Inference: An Integrated Approach* (London: Arnold, 1999).

الفصل الخامس

D. C. Montgomery, *Design and Analysis of Experiments* (New York: John Wiley and Sons, 2004).

L. Kish, *Survey Sampling* (New York: John Wiley and Sons, 1995).

الفصل السادس

G. E. P. Box, Robustness in the strategy of scientific model building, technical report, Madison Mathematics Research Center, Wisconsin University, 1979.

E. Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 2001).

A. Unwin, M. Theus, and H. Hofmann, *Graphics of Large Data Sets: Visualising a Million* (New York: Springer-Verlag, 2006).

مصادر الصور

All images credits go to © David Hand.